The Genetic Epidemiology and Omics of age-related cataract

Yonova-Doing, Ekaterina

Awarding institution:
King's College London

The Genetic Epidemiology and Omics of age-related cataract

Ekaterina Hristova Yonova

Thesis submitted to the University of London in fulfilment of the requirements for the degree of Doctor of Philosophy in Genetic Epidemiology

July 2018

Declaration

I, Ekaterina Hristova Yonova, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Ekaterina H. Yonova

Date 14th July 2018

Abstract

Age-related cataract is the leading cause of blindness worldwide and a major public health problem. The two most common types of age-related cataract are nuclear and cortical cataract. Nuclear and cortical cataracts are complex diseases, meaning both genetic and environmental factors play a role in their pathogenesis. In addition, it is increasingly clear that pathogenesis of age-related cataracts is intertwined with systemic factors related to ageing. Holistic approaches taking into account the systemic nature of the cataract risk factors are needed. High-throughput technologies have enabled detailed studies of the molecular changes and biological pathways underlying a variety of complex traits but their use in age-related cataract research is long overdue.

The current thesis describes the implementation of various “omics” platforms to answer questions related to genetic and environmental risks such as: 1) What are the genetic variants (both common and rare) that underlie the pathogenesis of the common forms of age-related cataract: nuclear and cortical cataract?; 2) Do nuclear and cortical share genetic risk factors?; 3) What are the effects of diet on cataract formation and progression and are these effects mediated through circulating metabolites or the gut microbiome?

The work presented in this thesis resulted in, among others: the identification of six new genetic loci for age-related nuclear cataract; determining that dietary vitamin C protects not only against cataract formation but also against cataract progression; and suggesting that gut bacteria may play a role in the pathogenesis of cataract. Omics data can be useful for elucidating the biological pathways underling cataract. However, various factors that lead to reduction in power to identify effects, such as mismatches between dates of phenotyping and sample collection, may limit their potential.

Acknowledgements

"Not everything that can be counted counts,
and not everything that counts can be counted"
William Bruce Cameron

I would like to thank all the volunteers from the TwinsUK and the International Cataract Genetics Consortium (ICGS) cohorts for providing their time and cooperation.

I would like to express my deep gratitude to my supervisors Prof. Christopher Hammond and Dr. Pirro Hysi. Dear Chris, thank you for your inspiring, supportive and patient mentorship and for encouraging me to pursue my research and teaching interests. Dear Pirro, thank you for being my statistics and methods guru and for challenging my perceptions.

I gratefully acknowledge my funding bodies, BBSRC and LiDO-dtp, which made my PhD possible.

I would also like to thank the members of my thesis committee Dr. Kerrin Small, Dr. Ananth Viswanathan and Dr. Valentina Cipriani for their support and helpful discussions. Kerrin, thank you also for giving me access to the eQTL data.

I would like to acknowledge the members of ICGS for their contribution to the work presented here, and especially: Dr. Ching-Yu Cheng for his general guidance and help with article writing; Prof. Caroline Klaver you for giving me free reign over the cataract slides; Prof. Sudha Iyengar for helping with the functional interpretation; Prof. Jei Jin Wong and Prof. Barbara Klein, for sharing their knowledge and experience and for encouraging me. Here, I would also like to thank the Rotterdam team: Gabriëlle Buitendijk, Virginie Verhoeven, Henriet Springelkamp, Adriana Iglesias and Milly Tedja; for making me feel home, and for all the sushi and cocktails.

Next, I would like to acknowledge the present- and ex-members of the Academic Ophthalmology team: Katie Williams and Cristina Venturini for all the support, ideas, unforgettable moments, and for your friendship; Abhishek Nag for being the D'Artagnan to our three musketeers; Zoë Forkin, Omar Maroo, Jelle Vehof, Mark Simcoe for the fruitful collaborations; Diana Kozareva for the support with the twin visits; and Maria Bell and Philippa Sacre for sorting out my bureaucratic mess. I would also like to thank my interns: Daniel Poon, Harry Rosen, Diego Casa, Antonioa de Candia, Eneh Jones, Eduard Bloch, Dan Wheeler, Claire Wong and Shehnaz Bazeer; for their work, input and for helping me develop my supervision skills.

I gratefully acknowledge all the members of the Department of Twin research and Genetic Epidemiology: Prof. Tim Spector for the opportunity of joining the department;

Genevieve Lachance for delivering data in the shortest possible time; Dr. Jonas Zierer and Dr. Cristina Menni for the access to the metabolome data; Matt Jackson, Dr. Claire Steves and Dr. Jordana Bell for the access to the microbiome data; Ruth Bowyer for the FFQ data; Dr. Alexander Alves for the help with the eQTL analysis and for the general discussions about science; Andrew Anastasiou for solving my IT problems; Johanna Honey for making the twin visits a pleasure; Craig Glastonbury and Tiphaine Martin for the gorgeous dinners during ASHG; Lisa Wolber for introducing me to the Brilliant club; and my fellow PhD students Leonie Roos, Idil Erte, Michelle Beaumont and Pei-Chien Tsai.

I would like to thank Professor Rebecca Oakey and Dr. Reiner Schulz at KCL, and Professor Mark Thomas at UCL, and their teams, for giving me the opportunity to work with them during my internship projects, and for expanding my scientific interests.

I would also like to thank everyone at the LiOD-dtp: Prof. Geriant Thomas for letting me have another project; Dr. QueeLim Ch'ng for all the helpful tips, Nadine Mogford for the amazing organisational skills and unforgettable events; Sarah MacDonald, Ben Bee and Jøsh Büllöck for the fascinating discussions.

I would also like to thank Dr. Joanna Howson from the Cardiovascular Epidemiology Unit in Cambridge for hiring me and guiding me through the murky waters of mitochondrial genetics.

Next, I would like to thank everyone at the Brilliant Club for sharing their knowledge on nurturing talent. Special thanks go to the immensely inspirational Mary Hanes for giving me the possibility to do my professional internship with the Brilliant club and for letting me stay on as a tutor, Rachael Cursons, Rajbir Hazelwood, and Andrew Fleming for tirelessly facilitating my teaching. Thanks go to the KCL widening participation team, and especially to Syreeta Allen, for giving me the opportunity to inspire so many young people. Thanks also go to my more than 200 students, but especially to the ones that were exceedingly driven and perceptive: Samantha, Sara, Amritha, Elias, Max, Anthony, Marcus and Adam.

I would also like to thank my friends: Bilyana Dicheva, Lisa Marr and Paolo Davanzo, Blanca Regina and Pierre Bouvier Patron, Bea Haut and Gavin O'Shea, Lara Torres and John Kannenberg, Steven Ball and Guli Silberstein; for being there for me and showing me that there is much more to life.

Finally, I would like to thank my family for their constant support. First and foremost, I would like to thank wholeheartedly my amazing husband Dr. Karel Doing for encouraging me to pursue my PhD, for always supporting me, discussing with me the ways of science, art, humanity and the universe in general and especially for cooking for me every day during my thesis writing ordeal. You are my best friend and you

always tell me that I can. Thanks also go to my mother, father and granny for being supportive and for believing in me.

This thesis is written in loving memory of the other PhDs in the family, Dr. Yordan Yonov, Dr. Hendrik Doing and Dr. Nora Huis in't Veld for their genes and inspiring lives.

Publications arising from this thesis

I. Articles and book chapters

Pascale Gerbault, Catherine Walker, Katherine Brown, **Ekaterina Yonova-Doing**, and Mark G. Thomas. The Evolution of Lactose Tolerance in Dairying Populations. The Oxford Handbook of the Archaeology of Diet; published online: August 2017

Williams KM, Hysi PG, **Yonova-Doing E**, Mahroo OA, Snieder H, Hammond CJ. Phenotypic and genotypic correlation between myopia and intelligence. Sci Rep. 2017 Apr 6;7:45977.

Mahroo OA, Williams KM, Hossain IT, **Yonova-Doing E**, Kozareva D, Yusuf A, Sheriff I, Oomerjee M, Soorma T, Hammond CJ. Do twins share the same dress code? Quantifying relative genetic and environmental contributions to subjective perceptions of "the dress" in a classical twin study. J Vis. 2017 Jan 1;17(1):29. doi: 10.1167/17.1.29.

Springelkamp H, Iglesias AI, Mishra A, Höhn R, Wojciechowski R, Khawaja AP, Nag A, Wang YX, Wang JJ, Cuellar-Partida G, Gibson J, Bailey JN, Vithana EN, Gharahkhani P, Boutin T, Ramdas WD, Zeller T, Luben RN, **Yonova-Doing E**, Viswanathan AC, Yazar S, Cree AJ, Haines JL, Koh JY, Souzeau E, Wilson JF, Amin N, Müller C, Venturini C, Kearns LS, Kang JH; NEIGHBORHOOD Consortium, Tham YC, Zhou T, van Leeuwen EM, Nickels S, Sanfilippo P, Liao J, van der Linde H, Zhao W, van Koolwijk LM, Zheng L, Rivadeneira F, Baskaran M, van der Lee SJ, Perera S, de Jong PT, Oostra BA, Uitterlinden AG, Fan Q, Hofman A, Tai ES, Vingerling JR, Sim X, Wolfs RC, Teo YY, Lemij HG, Khor CC, Willemsen R, Lackner KJ, Aung T, Jansonius NM, Montgomery G, Wild PS, Young TL, Burdon KP, Hysi PG, Pasquale LR, Wong TY, Klaver CC, Hewitt AW, Jonas JB, Mitchell P, Lotery AJ, Foster PJ, Vitart V, Pfeiffer N, Craig JE, Mackey DA, Hammond CJ, Wiggs JL, Cheng CY, van Duijn CM, MacGregor S. New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics. Hum Mol Genet. 2017 Jan 15;26(2):438-453.

Bloch E*, **Yonova-Doing E***, Jones-Odeh E, Williams KM, Kozareva D, Hammond CJ. Genetic and Environmental Factors Associated With the Ganglion Cell Complex in a Healthy Ageing British Cohort. JAMA Ophthalmol. 2017 Jan 1;135(1):31-38.

Jones-Odeh E*, **Yonova-Doing E***, Bloch E, Williams KM, Steves CJ, Hammond CJ. The correlation between cognitive performance and retinal nerve fibre layer thickness is largely explained by genetic factors. Sci Rep. 2016 Sep 28;6:34116.

Yonova-Doing E, Hammond CJ. Re: Datiles *et al.*: Longitudinal study of age-related cataract using dynamic light scattering: loss of α-crystallin leads to nuclear cataract development (Ophthalmology 2016;123:248-54). Ophthalmology. 2016 Aug; 123(8):e47-8.

Yonova-Doing E, Forkin ZA, Hysi PG, Williams KM, Spector TD, Gilbert CE, Hammond CJ. Genetic and Dietary Factors Influencing the Progression of Nuclear Cataract. Ophthalmology. 2016 Jun;123(6):1237-44.

Li Q, Wojciechowski R, Simpson CL, Hysi PG, Verhoeven VJ, Ikram MK, Höhn R, Vitart V, Hewitt AW, Oexle K, Mäkelä KM, MacGregor S, Pirastu M, Fan Q, Cheng CY, St Pourcain B, McMahon G, Kemp JP, Northstone K, Rahi JS, Cumberland PM, Martin NG, Sanfilippo PG, Lu Y, Wang YX, Hayward C, Polašek O, Campbell H, Bencic G, Wright AF, Wedenoja J, Zeller T, Schillert A, Mirshahi A, Lackner K, Yip SP, Yap MK, Ried JS, Gieger C, Murgia F, Wilson JF, Fleck B, Yazar S, Vingerling JR, Hofman A, Uitterlinden A, Rivadeneira F, Amin N, Karssen L, Oostr§§a BA, Zhou X, Teo YY, Tai ES, Vithana E, Barathi V, Zheng Y, Siantar RG, Neelam K, Shin Y, Lam J, **Yonova-Doing E**, Venturini C, Hosseini SM, Wong HS, Lehtimäki T, Kähönen M, Raitakari O, Timpson NJ, Evans DM, Khor CC, Aung T, Young TL, Mitchell P, Klein B, van Duijn CM, Meitinger T, Jonas JB, Baird PN, Mackey DA, Wong TY, Saw SM, Pärssinen O, Stambolian D, Hammond CJ, Klaver CC, Williams C, Paterson AD, Bailey-Wilson JE, Guggenheim JA; CREAM Consortium. Genome-wide association study for refractive astigmatism reveals genetic co-determination with spherical equivalent refractive error: the CREAM consortium. Hum Genet. 2015 Feb;134(2):131-46.

Nag A, Venturini C, Small KS; Aung T, Cheng CY, Fleck BW, Gibson J, Hewitt AW, Hofman A, Höhn R, Jonas JB, Khor CC, Klaver CC, Lemij HG, Liao J, Lotery AJ, Lu Y, Macgregor S, Mitchell P, Ramdas WD, Springelkamp H, Tai ES, Teo YY, Uitterlinden AG, van Duijn CM, van Koolwijk LM, Vingerling JR, Vitart V, Vithana E, Wang JJ, Williams KM, Wojciechowski R, Wong TY, Xu L, **Yonova-Doing E**, Tanja Z., Young TL, Viswanathan AC, Mackey DA, Hysi PG, Hammond C. A genome-wide association study of intra-ocular pressure suggests a novel association in the gene FAM125B in the TwinsUK cohort. Hum Mol Genet. 2014 Jun 15;23(12):3343-8.

II. Selected presentations

Yonova, E *et al.* Associations between the gut microbiome, a healthy diet, and age-related nuclear cataract. **ARVO 2017 poster**;

Shehnaz Bazeer, **Ekaterina Yonova**, Katie M. Williams, Diana Kozareva, Mark Simcoe, Christopher J. Hammond. Cognitive, visual and genetic factors influencing the Useful Field of View (UFOV) test: a twin study; **ARVO 2017 poster**;

Claire L. Wong, Ekaterina Yonova, Katie M. Williams, Mark J. Simcoe, Diana Kozareva, Christopher J. Hammond. Keratoconus Endophenotypes in a Healthy Ageing Population, **ARVO 2017 poster;**

Katie M. Williams, **Ekaterina Yonova**, Mark J. Simcoe, Jelle Vehof, Pirro G. Hysi, Christopher J. Hammond. Early life factors for myopia in the Twins Early Development Study (TEDS); **ARVO 2017 paper**;

Yonova E *et al.*, Common and rare variants of congenital cataract genes and age-related nuclear cataract: ICGC; **ASHG 2016 poster**;

Yonova E *et al*. / **Yonova-Doing E** *et al*. Genome-wide multi-ethnic meta-analyses identify new loci associated with age-related nuclear cataract; **ARVO 2016 paper; GMM 2016 poster**, **Quantitative genetics 2016 poster. LiDO retreat 2016, poster.**

Wanting Zhao, Rob P. Igo, **Ekaterina Yonova-Doing,** Gyungah Jun, Caroline C. Klaver, Barbara E. Klein, Christopher J. Hammond, Ching-Yu Cheng, Jie Jin Wang. Genome-wide trans-ancestry meta-analysis identifies new susceptibility loci for age-related cortical cataract; **ARVO 2016 poster**;

Edward Blochm, **Ekaterina H. Yonova,** Eneh JONES-ODEH, Katie Williams, Diana Kozareva, Christopher J. Hammond. Genetic and environmental factors influencing the ganglion cell complex in a healthy ageing British cohort, **ARVO 2016 paper;**

Eneh Jones-Odeh, **Ekaterina H. Yonova,** Edward Bloch, Katie Williams, Claire Steves, Christopher J. Hammond. Is the relationship between retinal nerve fibre layer thickness and cognitive performance explained by genetic or environmental factors? A twin study;

ARVO 2016 poster;

Yonova E., *et al.* Genome-wide association study of nuclear cataract finds suggestive association with a common variant in *TRPM3*; **ARVO 2015 paper;**

Katie M. Williams, Cristina Venturini, **Ekaterina H. Yonova**, Pirro G. Hysi, Robert Plomin Christopher J. Hammond. Evidence for shared genetic factors between myopia and intelligence in the Twins Early Development Study (TEDS); **ARVO 2015 paper**

Yonova-Doing E. *et al.* Genetic and environmental factors in age-related nuclear cataract; **Young Researcher Vision Camp 2015 poster;**

Yonova-Doing E *et al.* PhD thesis in 3 minutes (talk); **UKEgg 2016; LiDO 2015**

Table of Contents

LIST OF FIGURES

LIST OF TABLES

KEY OF ABBREVIATIONS

A	Additive genetics effect
AIC	Akaike information criterion
ARC	Age-related cataracts
AREDS	Age-Related Eye Diseases Study, or the cataract grading system developed by this study
aSUM	Data-adaptive sum test
BAES	Barbados Eye Study
BDES	Beaver Dam Eye Study
BES	Beijing Eye Study
BMES	Blue Mountains Eye Study
BMI	Body mass index
BMP	Bone morphogenic protein
C	shared environment
CARESD	Carotenoids in Age-Related Eye Disease Study
CAST	Cohort allelic sums test
CI	Confidence Interval
CHBES	Chesapeake Bay Eye Study
CMC	Combined and Multivariate Collapsing
CNDS	Central nuclear dip score
CVD	Cardiovascular diseases
DTR	Department of Twin Research and Genetic Epidemiology
DZ	Dizygotic twin
eMERGE	Electronic medical records and genomics network
E	unique environment and measurement error
EPACTS	Efficient and Parallelizable Association Container Toolbox
EPIC	European Prospective Investigation into Cancer and Nutrition
EREC	Estimated regression coefficient
EST	Expressed sequence tag
FDR	False discovery rate
FFQ	Food frequency questionnaire
FPKM	Fragments per kilobase of transcript per million mapped reads

GC	Gas chromatography
GSH	Glutathione
GWAS	Genome-wide association studies
HATS	Healthy Ageing in Twins Study
HDL	High density lipoprotein
HEI	Healthy Eating Index
HMW	High molecular weight
HRC	Haplotype reference consortium
HRT	Hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
indel	Insertion deletion polymorphism
IQR	Intraquartile range
KBAC	Kernel-based adaptive cluster
LC	Liquid chromatography
LD	Linkage disequlibrium
LEC	Lens epithelium cells
LFRGV	Low-frequency and rare genetic variants
lncRNA	Long non-coding RNA
LOCCS	Lens Opacities Case Control Study
MAF	Minor allele frequency
MDS	Mediterranean diet score
miRNA	MicroRNA
MS	Mass spectrometry
MVIP	Melbourne Visual Impairment Project
MZ	Monozygotic twin
N	number of (may refer to individuals/samples or SNPs)
NMR	Nuclear magnetic resonance
NurHS	Nurses Health Study
OR	Odds ratio
OTU	Operational taxonomy unit
PCA	Principal component analysis
PHS	Physicians Health Study

PLC	Prospective lens epithelium cells
PSC	Posterior subcapsular cataract
POLA	Pathologies Oculaires Liées á l'Age
PUFAs	Polyunsaturated fatty acids
QC	Quality control
RAAB	Rapid assessment of avoidable blindness
RES	Reykjavik Eye Study
ROS	Reactive oxygen species
RSI-III	Rotterdam Study I - follow-up 3
SD	Standard deviation
SEEP	Salisbury Eye Evaluation Project
SiMES	Singapore Malay Eye Study
SKAT	SNP-set (Sequence) Kernel Association Test
SNP	single nucleotide polymorphism
SSU	Sum of squared scores
SUN	Seguimiento University of Navarra
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TF	Transcription factor
UV	Ultraviolet radiation
V	Variance
VT	Variable threshold test
WHO	World Health Organisation
WNT	Wingless integrator proteins
1000G	1000 Genomes Project

Chapter 1. Introduction

1.1. Overview

Age-related cataract (ARC) is an ophthalmic condition that arises from the interplay of genetic and environmental factors. The work discussed within this thesis aims to elucidate some of the genetic factors underlying ARC and, through the use of high-throughput technologies, the mechanisms behind the association between ARC and diet, one of ARCs environmental determinants.

In order to show the importance of studying ARC, I will start this chapter by discussing the global impact of cataract on eye health. Next, I will define what ARC is and how it is measured. In order to understand the biological mechanisms underlying ARC formation, it is helpful to understand the anatomical, physiological and developmental characteristics of the human lens, the part of the eye that cataract affects. After addressing what we know about cataract biology, I will focus on the genetic and environmental risk factors for ARC development and progression.

1.2. Cataract as a public health problem

Age-related cataract is a major public health problem and a leading cause of blindness worldwide[1]. According to the latest summary statistics, the number of blind due to cataract worldwide rose from 12.3 million in 1990 to 20 million in 2010[1,2]. Cataract related blindness is not distributed evenly around the world as illustrated by data collected between the late 1990 and 2010: in North America* and West Europe cataract accounts for 12.7% and 13.8% of blindness respectively, while in East Sub-Saharan Africa cataract accounts for 36.7% of blindness and in Southeast Asia for 42.0%[3-5]. Cataract-related blindness can be avoided by surgically removing the lens and replacing the affected lens with an artificial one. A valuable resource of information of cataract related blindness and severe visual impairment after the year 2010 comes from the rapid assessment of avoidable blindness (RAAB) studies[6]. RAAB uses cluster-based randomised door-to-door survey and visual acuity testing strategy to assess the prevalence and causes of blindness in a survey area, divided into segments with sample size of 50 participants over the age of 50[6]. The RAAB is ideally designed for obtaining a total sample size of between 2000 and 5000 individuals. However, the studies provide predominantly raw prevalence which makes comparison between

* high income

countries difficult. The findings of those studies are summarised in **Table 1.1**.

Although overall the worldwide prevalence of cataract-related blindness is decreasing and the rates of cataract surgery are increasing[7], there are still a lot of areas with high prevalence of cataract related blindness. And given an increasing elderly population in many countries the above mentioned positive changes do not happen fast enough. Moreover, lack of ophthalmologists in developing countries and the significant costs of treatment and care in the industrialised countries pose significant issues to the national healthcare systems[7]. In the USA alone it is estimated that direct medical costs for cataract treatment are an estimated $6.8 billion per year[8]. In UK for the year 2010 the estimated cost of cataract treatment (referral and surgical treatment) and ongoing social and personal care for those who are partially sighted or blind due to cataract was over £995 million[9]. Ninety percent of visually impaired people live in developing countries where such high spending levels are not possible. For example, in Nigeria the median direct cost of cataract surgery equals more than 50 days income for 70% of the local population[10].

Table 1.1. Summary of results from the rapid assessment of avoidable blindness studies

		cataract-related blindness		cataract-related visual impairment	
Region	**N**	**Lowest**	**Highest**	**Lowest**	**Highest**
AFR(N) /ME	4	41.0 (Saudi Arabia)	58.6 (Saudi Arabia)	45.0 (Saudi Arabia)	71.0 (Saudi Arabia)
AFR(S)	3	39.8 (Zambia)	55.0 (Burundi)	43.0 (Burundi)	82.7 (Madagascar)
AMR	1	63.0 (Mexico)		68.0 (Mexico)	
EAS/ SEA	4	51.4 (Maldives)	81.8 (Cambodia)	59.5 (China)	94.1 (Cambodia)
EUR	2	21.2 (Hungary)	58.2 (Moldova)	35.3 (Hungary)	70.2 (Moldova)
SAS	7	56.0 (India)	74.6 (India)	51.4 (India)	82.4 (India)

This table summarises the results from the rapid assessment of avoidable blindness studies that were published after 2010. For each world region the lowest and highest raw prevalences (%) measured are presented, followed by the country where this was measured. AFR(N)/ME – Africa (North) and the Middle East; AFR(S) – Africa (Subsaharan); AMR – Latin America; EAS/ SEA – East and South-East Asia; EUR – Europe; ME - SAS -South Asia. N – number of studies conducted per region. The figure for this table come from the studies cited in the references list with numbers:11-24.

Cataract will likely continue to be a major public health problem, even as we are approaching the end of the World Health Organisation (WHO) VISION 2020 initiative,

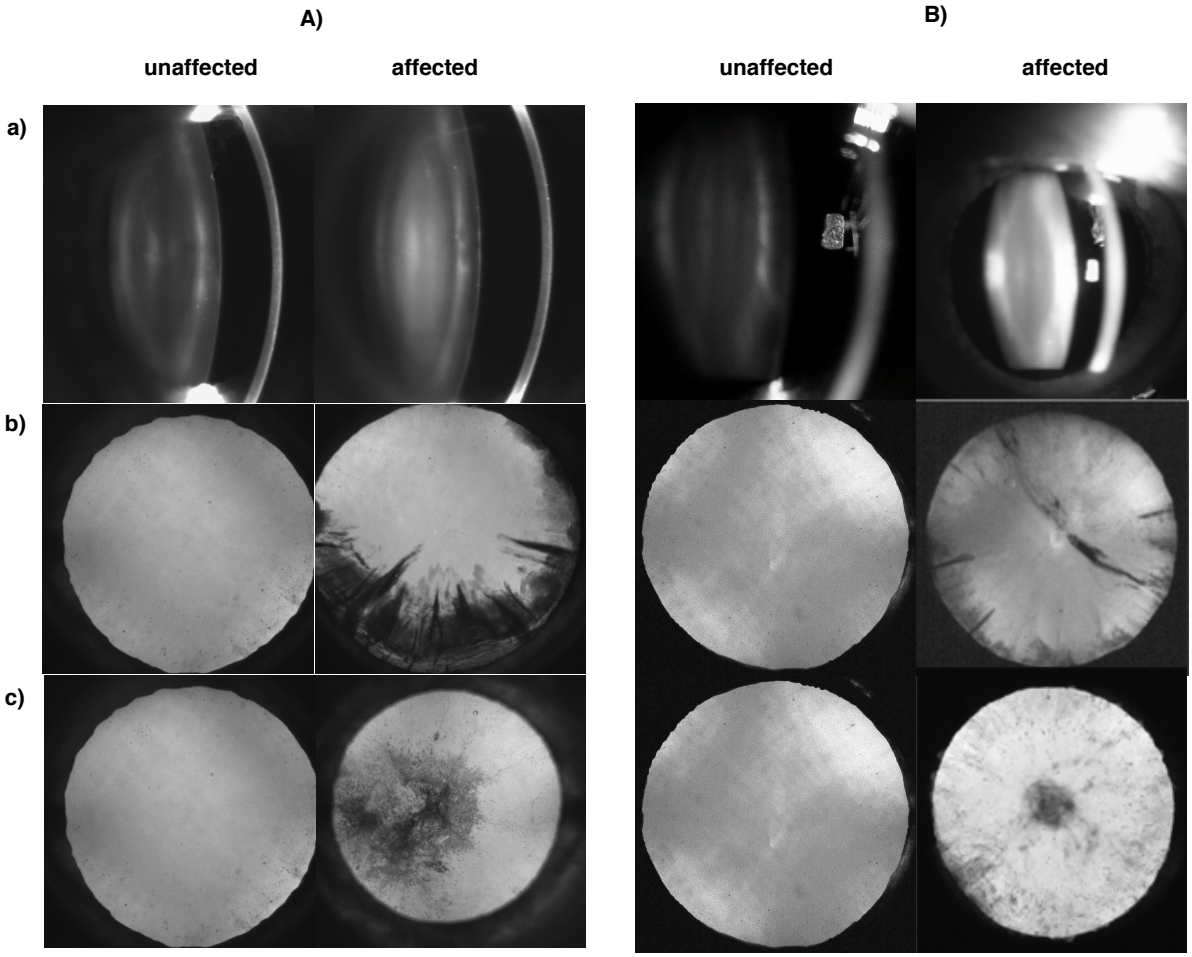
which targeted the eradication of cataract-related blindness. There are several reasons for this phenomenon. The first two are the higher life-expectancy worldwide and the sustained population growth in areas with high cataract prevalence. There is also an increase in prevalence of several of the risk factors for developing cataract, such as diabetes and obesity. Last but not least, there is still a limited access to treatment in poor countries, particularly sub-Saharan Africa. In this situation, cataract prevention or non-surgical interventions may prove to be a better way forwards.

1.3. Age-related cataract: definitions and assessment

ARC is defined as slow and progressive loss of transparency (opacification) within the clear lens of the eye as the eye ages. The opacity interferes with the passage of light causing impaired visual acuity and glare due to scatter. ARC is a complex trait caused by the interplay of genetic and environmental factors[25-27]. ARC generally affects people above the age of 50. The lens shows a number of zones of optical discontinuity in which transparent regions are separated by bands of increased light-scatter and reflection[28]. The age-related opacification of the lens has different morphological presentation at the different zones, resulting in the three most common types of ARC: nuclear cataract, cortical cataract, and posterior subcapsular cataract (PSC) (**Figure 1.1.**). Nuclear cataract affects the fibres at the centre of the lens (the fetal nucleus) and presents as uniform clouding, and coloration of the lens from grey-white and light-blue (opalescence) in its early stages to yellow-brown (brunescence) and deep blackish-brown (cataracta nigra) in the advanced stages[28,29]. Nuclear cataract affects the central vision and is the most common indication for cataract surgery.

Cortical cataract affects the fibres in the lens cortex and presents as radial spokes[28-30]. The spokes start in inferonasal quadrant of the lens as a small numbers of opaque fibre bundles that grow and spread both along individual fibres and also to adjacent fibres[28,31]. This results in centripetal and circumferential spoke growth at the equator[28,31]. Cortical cataract affects the peripheral vision causing glare and dazzle, particularly in dim light when the pupil is large, but, in later stages, it in can also affect the central vision. PSC cataract affects the posterior cortex just under the posterior capsule and forms web-like structures, thus also affecting central vision[28-30]. PSC cataract is due to defective fibre production by the epithelium that results in migration of metaplastic cells from the lens equator to the posterior pole and in failure of lens growth[28].

Figure 1.1: Types of age-related cataract



In this figure black and white lens images from participants in the TwinsUK (A) and the Rotterdam study (B) cohorts are presented. The images exemplify the tree types of age-related cataract: nuclear cataract (a), cortical cataract (b) and PSC cataract (c).For comparison lens images from unaffected eyes are also included.

PSC has a disproportionate effect on vision as it develops centrally at the back of the lens, at the nodal point where rays of light converge. The three types of cataract can present alone or in combination and they are typically bilateral but commonly asymmetrical[30].

Cataracts are detected, either by direct visualisation or by using optical equipment, with or without lens photography, and one of the following systems can be employed: ophthalmoscopy (direct/retroilllumination), photography (Scheimpflug or retro-illumination); optical degradation of projected retinal images; fluorimetry; and laser quasielastic light scattering[32]. The second and third are used most in research and will be discussed here. The slit lamp contains a high-intensity light source that can be focused to shine a thin sheet of light into the eye. The reflected image is then amplified

with a biomicroscope. All types of cataract can be detected using a slit lamp but only single section of the lens is visible at any one time which, when used in combination with photography, can be limiting. During retro-illumination the lens and photographic camera are parallel and light beam is shone straight in the eye, so that the lens is illuminated only by light reflected from the back of the eye. The reflected light is tilted by a mirror system and captured by a camera. This technique is particularly good for detection of cortical and PSC cataracts. Scheimpflug photography is used for detecting nuclear cataracts, as it provides depth of focus across the whole lens. A light beam hits the ocular lens straight in the visual axis, while the camera records the image under an oblique angle allowing depth of focus through the lens.

Table 1.2. Characteristics of existing subjective cataract grading systems

Grading System	Score types	N of SLP/SLD	SLP/SLD
AREDS	decimal	3x cataract type	AREDS stnadard photos
JAP-CCESG	integer	2x NC and CC; 1x PSC	JAP-CCESG standard photos
LOCS II	decimal	4x NC and PSC; 5x CC	LOCS II lens photos
LOCS III	decimal	6x cataract type	LOCS III lens photos
Oxford$	integer	10x NC, 5x CC and PSC	Diagrams only; NC: brunescence/white scatter; CC: 1/5 pie segments; PSC: concentric segments;
Oxford&	decimal		
WHO	integer	4 x cataract type	NC photos; diagrams: CC 1/8 pie segments; PSC ⌀
Wilmer	integer	4 x NC and CC; ⌀ PSC	NC photos, a CC 1/8 pie diagram, PSC dimensions (slit lamp calibration beams)
Wisconsin$	integer	4 x cataract type	Wisconsin standard lens photos
Wisconsin&	decimal		

SLT/SLD — standard lens photos/diagrams used for comparison; N — number of SLP/SLD used for comparison; AREDS — Age-Related Eye Diseases Study grading system[33]; JAP-CCESG — Japanese Cooperative Cataract Epidemiology Study Group classification system[34]; LOCS II/III — Lens Opacities Classification System versions II and III[35,36]; Oxford — Oxford Clinical Cataract Classification and Grading System[37]; Wilmer — Wilmer Grading System[38]; WHO — Simplified Cataract Grading System (http://www.who.int/iris/handle/10665/67221); Wisconsin — Wisconsin System for Classification of Cataracts from Photographs[39]; $ — original; & — modified; NC - nuclear cataract; CC - cortical cataract.

Once the lens is visible or lens photographs are obtained, cataract severity is usually graded by subjective comparison to a standard set of lens photographs.

Objective densitometry systems exist but are rarely used. The main characteristics of the existing subjective cataract grading systems are presented below (**Table 1.2.**). The various grading systems produce different prevalence measures but studies, where two or more grading systems were compared, showed that calibration between systems is easily achievable and that grading systems produce results that are to some extent comparable[40-42].

1.4. Anatomy and embryonic development of the lens

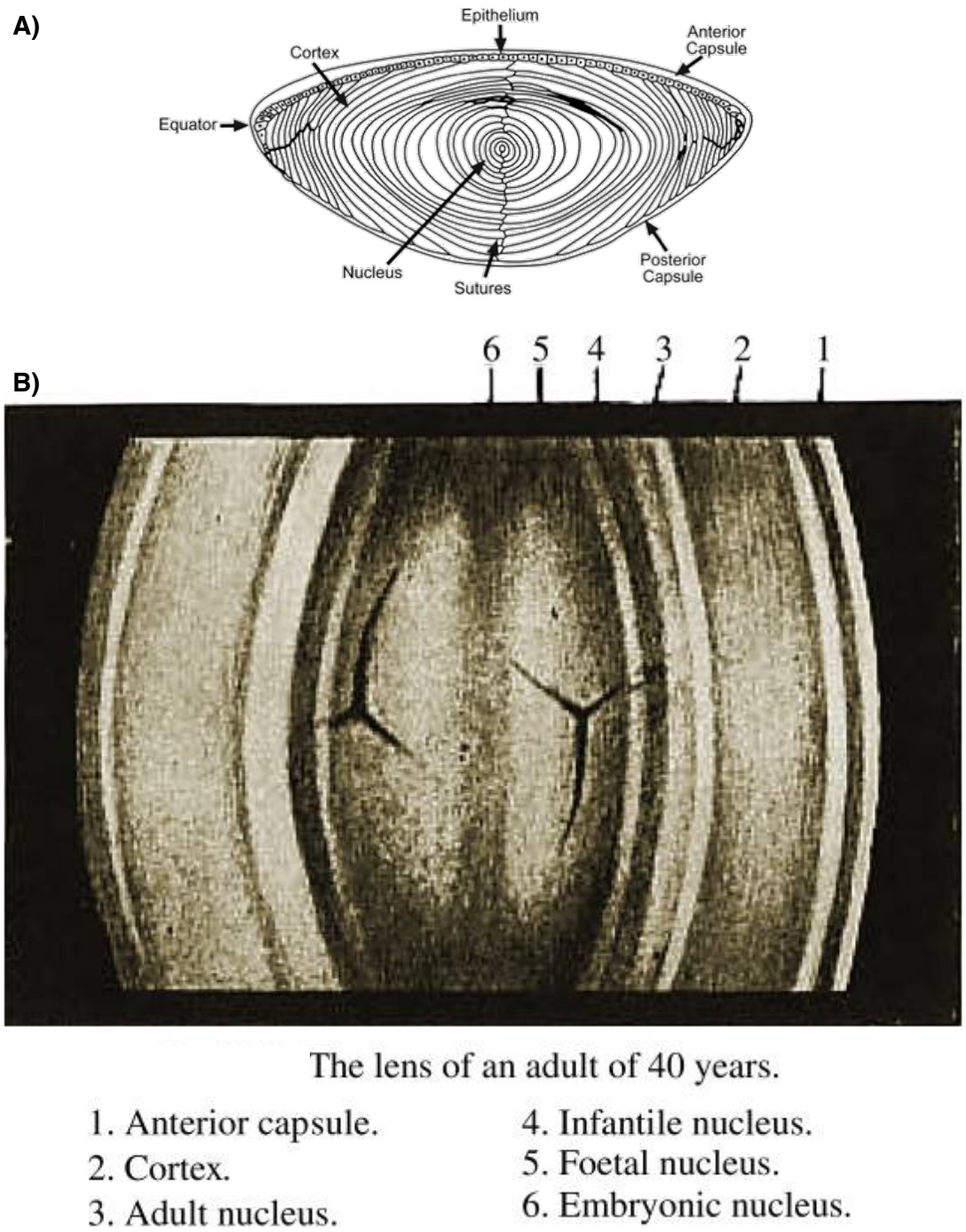
The human lens is a transparent, avascular, non-innervated, biconvex structure (~10 mm in diameter) in the anterior segment of the eye, which continues grow throughout adulthood[43]. Together with the cornea, the lens helps focus the light on the retina and by altering its shape through the process of accommodation it can fine-tune the eye's focus. The human lens has three histological compartments: lens capsule, lens epithelium, and lens fibres (**Figure 1.2**)[43].

The lens capsule is a basement membrane, rich in collagen type-IV, laminin and fibronectin[43]. It has structural and signalling functions, and allows the passage of small molecules both into and out of the lens[43,44]. The lens epithelium is a simple cuboidal epithelium that regulates the lens osmolarity and the transport of molecules from the aqueous humor, and gives rise to the lens fibres. The fibres cells form the bulk of the lens mass and are long (4µm diameter and up to 12 mm length), thin, transparent cells that, when mature, are devoid of organelles and nuclei and connected to each other by adherens and gap junctions[45]. The lens fibres stretch lengthwise from the posterior to the anterior poles forming honeycomb-like onion-layered structures (**Figure 1.2**). They are continuously produced; until early childhood the lens grows asymptotically and after that point the growth becomes linear[46].

The healthy human lens is transparent to visible light and nearly transparent in the long-wavelength region but absorbs strongly in the UV and short-wavelength regions[47]. The lens transparency is due to the lack of organelles and to its special protein and membrane organisation. The lens has the highest protein content of any other tissue in the body (about 38% of the total mass) and up to 90% of which are crystallins: small-sized (<10nm in diameter) closely-packed water-soluble proteins[28,48]. Lens proteins also have a long half-life and are characterised by long-term stability, which is due to the inhibition of degradative enzymes and a shift that cause ubiquitin to bind to proteins under stress and, protect them from, rather than tagging them for, degradation[49]. The minimisation of scatter is further enhanced by the tight packing of

lens fibre cells and changes in the concentrations of lipids in the membrane: high cholesterol ratios; low percentage of phosphatidylcholine; higher percentage of phosphatidylglycerol, lysophosphatidylglycerol, and gangliosides[50-53]. The lack of organelles results in anaerobic glycolysis being the only way ATP is produced in the lens fibres, with only 30% of ATP derived from the aerobic oxidation in the lens epithelium[43].

Figure 1.2. The structure of the adult human lens



The structure of the adult human lens is illustrated using a schematic representation of the

lenticular histological components (A); and a slit lamp biomicroscopy image (B) revealing the lens zones of optical discontinuity. Figure 1.2.A is adopted from Harding *et al*. 2002[54], while figure 1.2.B. is adopted from *et al*. Augusteyn[55] 2010.

Given the unique anatomical characteristics of human lens, it is perhaps not surprising that Mendelian mutations in genes encoding for structural lens proteins have been implicated in the formation of congenital cataracts in both animal models and in humans. Such genes include: crystallins (CRYA, CRYB and CRYG families), collagens (eg. *COL11A1* and *COL4A1*), aquaporins (eg. major intrinsic lens fibre protein (*MIP*) and *APQ1*), gap-junction proteins (eh. *GJA3 and GJA8*)[56-64].

To understand the anatomy and function of the lens, it is necessary to follow its embryonic development. Although disruption of normal lens development may lead to congenital cataract formation, it is unclear to what extent developmental events impact lens ageing. In human, lens development starts at embryonic day 22 with the formation of the lens placode from surface ectoderm cells that are adjacent to neural plate[65]. The placode starts envaginating and forms the lens vesicle[65]. Synthesis of crystallins, the structural lens proteins, starts at about embryonic day 30 and by embryonic day 33 the lens is fully separated from the optic vesicle[66]. After this point the lens cells start elongating towards the anterior of the lens vesicle and thus form the embryonic nucleus[65]. The cells of the anterior portion of the lens vesicle give rise to the lens epithelium, which in turn gives rise to additional secondary fibres located toward the equatorial region of the lens. Studies in various vertebrate animal models have shown that, from cellular and molecular prospective, lens morphogenesis can be divided into four phases: formation of lens placode from prospective lens ectoderm (PLE); invagination of the lens placode and formation the lens pit and lens vesicle; primary lens fibre cell differentiation; and fibre cell maturation that includes cytoskeletal modifications and organelle degradation[67]. Various genes do play a role at these stages. The most important of those, and their role in congenital cataract formation, will be discussed below.

During the first phase, the lens epithelium is established from multipotent placodal precursors which are induced to form the lens placode. PLEs arise from epithelial cells under the controls of bone morphogenetic proteins (BMPs), fibroblast growth factors (FGF) and Wingless integrator proteins (WNT)/β-catenin signaling[68]. The transition from PLE to placodal precursors cells and the establishing of lens identity starts with the action of two lysine acetyltransferases: CBP and p300[69]. They modify core histone proteins thereby initiating the expression of key lens TFs, such as Pax6, Six3 and Sox2-Oct1, that play role in the formation of the lens placode[67,69]. The

induction of the lens placode is dependent on signals coming from different sources and all those signals converge on Pax6 (**Figure 1.3.**), as discussed below. At gastrula stages, sustained exposure to autocrine bone morphogenetic proteins (Bmp2, Bmp4 and Bmp7), a subset of the TGF-β superfamily of growth factors, is required and sufficient to induce lens placodal cells[70,71]. In mice, targeted deletions of different components of the BMP pathway lead to disturbed lens formation[72]. At slightly later stages, BMP activity has been suggested to be required for crystallin expression in primary lens fibre cells and BMP receptor signalling also regulates both cell survival and proliferation during lens placode formation[73]. At this placode stage, paracrine signals from the prospective retina and from retinoic acid are also important[67]. The retinoic acid signalling also continues to play a role during the next phase of development when the lens placode invaginates to form the lens pit and lens vesicle[67]. Inhibition of cellular retinal binding protein (CRBP)-1 in mouse led to failure of lens placode invagination, and inactivation of retinaldehyde dehydrogenase led to disrupted lens vesicle formation[74].

The most important players at the first phase, however, is probably Pax6 (**Figure 1.3.**). Pax6 regulates a lot of different key lens development TFs, such as Six3, Sox2-Oct1, Sox11[75-77]. The functions of those will briefly be discussed below. Furthermore, Pax6 inhibits the WNT/β-catenin signalling which is a prerequisite for lens placode formation[78]. Apart from regulating the expression of a variety of different transcription factors, Pax6 can regulate its own activity using a positive feedback loop and is further regulated by a whole array of upstream proteins from different signalling pathways: Pygo2 (Wnt-independent lens induction); Raldh2, Rx (retinoic acid signalling); Hes1, Lhx2 (Notch signalling); Nf1 (ERK signalling); Mab21l2, Bmp4, and Bmp7(TGF-β); Fgfs (Fgf singling); Meis1 and Pknox1 (PI3K-Akt-mTOR)[79].

The effects of Pax6 mutations on eye development have been extensively studied. Pax6 knock-outs do not develop ocular structures and die in the neonatal period[80]. Heterozygote mice present with: microphthalmia (small eye), delay in lens placode formation, a smaller than normal lens, vacuolated lens fibres, and, importantly, cataracts in adults[81,82]. Humans with mutations in *PAX6* present with microphthalmia, aniridia (lack of iris) and aphakia (lack of lens), with only one study reporting congenital cataract[83,84].

An important target and co-regulator of Pax6 is Six3. In mice and chicks, *Six3* is initially expressed in PLE and the ectoderm overlying the optic vesicles; followed by expression in the lens placode and is, during differentiation, localised to the lens epithelium[85-87]. In humans the predominant phenotype caused by mutations in this gene is lack of forebrain development, while in animals, ectopic eyes and ectopic

lenses have been reported[88].

Figure 1.3. Pax6 extracellular (A) and intracellular (B) signalling in vertebrates

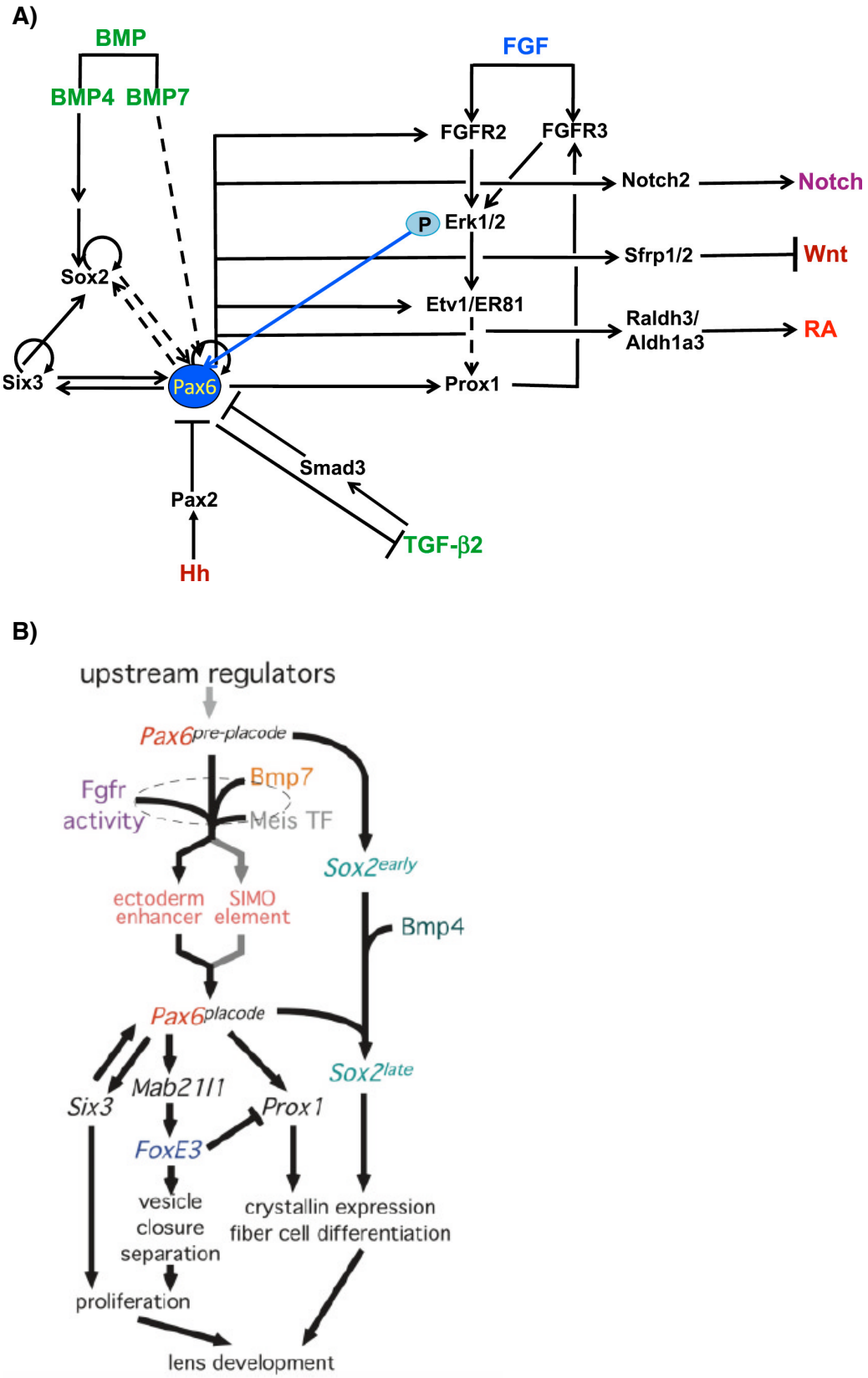


Figure A), adopted from Cveki et al (2017)[77], illustrates how Pax6 directly regulates the expression of components of six signal transduction pathways (FGF, Hedgehog-Hh, Notch, retinoic acid-RA, TGF-β/BMP and Wnt) which are upstream of Pax6. Figure B), adopted from Lang (2004)[88], shows some of the downstream targets of Pax6.

A group of key lens development TFs, under Pax6 regulation, is the Sox gene family of TFs (eg. Sox1, Sox2, Sox3, and Sox11). Sox genes are involved in the developmental processes leading to cell fate decisions and, dependent on tissue-specific partnering, activate a trans network of genes, including crystallins (murine γ F-crystallin, chicken δ1-crystallin genes)[89-91]. In absence of both Pax6 and Sox2 lens development is not initiated[92]. Similarly, in absence of Bmp4 which is upstream from both *Pax6* and *Sox2*, *Sox2* expression is not up-regulated, and the lens does not form[93]. Furthermore, it has been shown that *Pax6* expression is stabilised and maintained by Sox2 activity and that Oct-1, a Sox2 co-factor, is needed for this process[94,95]. Interestingly, Oct-1 has been shown to be an oxidative and metabolic stress sensor[96]. Pax6 and Sox2 together activate the transcription of lens specific MAFs (*L-Maf* and *c-Maf*) and which in terns bind to α A-crystallin and γ F-crystallin promoters respectively[97-99]. Disruption of the mouse c-mad gene led to failure of lens differentiation and elongation and a small hollow lenses phenotype[100]. In the lens, *Sox1* and *Sox3* have the same expression patterns as *Sox2* and they overlap with *Sox2* in terms of crystalline expression[90]. Mutations in all three genes caused microphthalmia in animal models and in humans but only *Sox1* mutant mice developed cataract[101,102]. Finally, Sox11 is important for lens placode invagination[79].

In vertebrates, including humans, *FoxE3* is another important member of the Pax6 regulatory network. In the mouse lens, *FoxE3* expression is first detected at E9.5 but following lens vesicle formation, *FoxE3* expression ceases in lens fibre cells but persists in the anterior epithelial cells[103,104]. FoxE3 is essential to lens vesicle closure and subsequent separation of the lens vesicle from the ectoderm, although the underlying mechanisms have not been extensively studied[105]. It is also involved in the maintenance of central lens epithelial cell proliferation via the upregulation of *Mki67*, *Pcna* and *Pdgfra* expression and the maintenance of equatorial lens epithelial cell phenotype via the exclusion of Prox1-induced Cdkn1c expression[103]. This prevents early cell cycle exit and premature lens epithelial cell differentiation into lens fibre cells[105]. Homozygous loss of function mutation in the human *FOXE3* present with a fusion between lens and cornea and an arrest of development of anterior epithelial and lens fibre cells, and congenital cataract[106,107].

Pax6 also regulates the expression of *Prox1, Ptx3, Otx1*, *Otx2*, *Creb-2, AP-2a,*

Eya1, Eya2. Prox1 seems to be driving γ D-crystallin expression and lens fibre differentiation while Ptx3 is involved in maintenance of the anterior epithelium of the lens[79]. Mutations in *Otx1/2* cause defects with varying severity including aphakia, microphthalmia and aniridia[108,109]. Furthermore, as implied by low levels of α A-crystallin in *Otx* knockout lens remnants, mutations in *Otx1/2* cause delay in or failure of lens development progression[110]. *Creb-2* is as an essential survival factor for anterior epithelial cells[111]. In mutant *Creb-2* mice, anterior epithelial cells undergo p53-dependent apoptosis at approximately stage E14.5, followed by resorption of the developing lens[111]. AP-2a is important for lens vesicle separation[79].It has also been shown to bind to *Pitx3* regulatory regions and to the promoter of MIP[112,113]. Both *Eya1* and *Eya2* transcripts have been detected in the developing lens and are prosed to play roles in lens placode development[114].

Following the first two phases of lens development there is a shift in regulatory pathways that lead to the initiation of the differentiation of primary lens fibres and epithelium cells[67]. Although the BMP, FGF and Notch pathways still play important roles in this phase, a variety of other regulatory pathways get activated. For example, vitreal factors such as insulin-like growth factors, platelet-derived growth factor and epidermal growth factor have been shown to potentiate lens cell proliferation, and later cell differentiation[115]. Another example is the members of B-cell lymphoma 2 (Bcl-2) family. Chick embryo experiments have shown that pro-apoptotic members of Bcl-2 are up-regulated in the lens and anti-apoptotic members are down-regulated at the time that organelle loss commences[116]. Mice exhibiting lens-specific Bcl-2 over-expression showed abnormal lens fibre cell differentiation, inducing cataracts, microphakia, vacuolization, lens fibre cell disorganisation and inhibition of lens fibre cell denucleation[117]. The expression of another important apoptosis regulator, p53 has detected in lens epithelium cels and in the lens fibres, and temporally distinct patterns of p53-dependent apoptosis have also been identified during mouse lens development[118,119]. The amount of p-53 signalling may be under the control of phosphatidylinositol 3-kinase[120]. Finally, tumour necrosis factor (TNF) related molecules and their signalling components have been shown to be expressed in the developing lens suggesting that TNF-related signalling (potentially TNFα-mediated cell signalling cascades) might be involved in regulating organelle loss in lens fibre cells and/or lens fibre cell differentiation[121]. Last but not least, cell–cell signalling plays an important role in fibre differentiation as shown by studies of human cataracts where mutations in EPHA2, encoding a receptor tyrosine kinase, were associated with autosomal dominant congenital cataracts and age-related cortical cataracts[122,123]. Furthermore, analyses of wild-type and Efna5–/– lenses indicated that Ephrin–A5/

EphA2 signalling regulates N-cadherin/β-catenin interactions, and thus adhesion and structure of the fibre cells[124].

1.5. The biology of age-related cataract formation

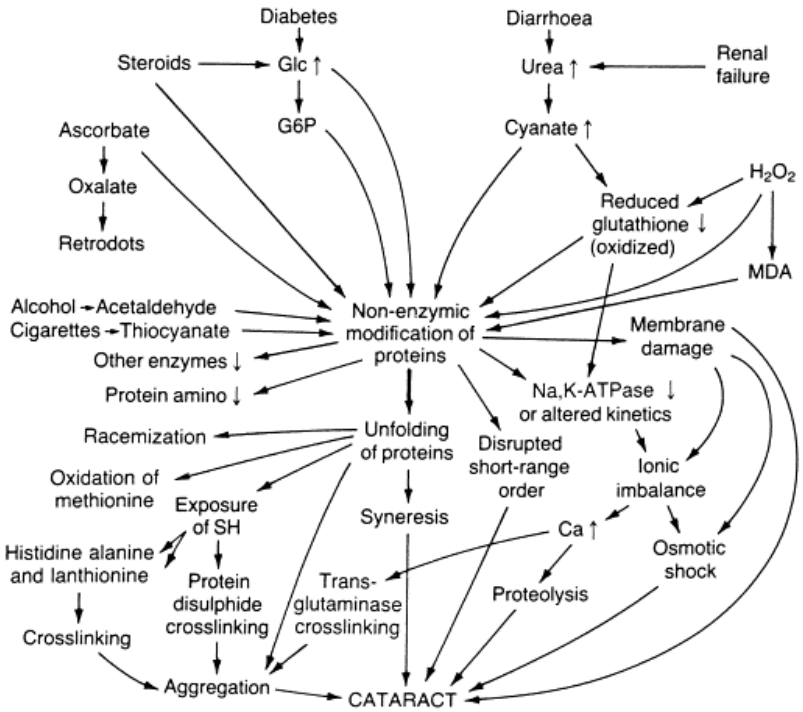
Understanding the biology of ARC is complicated by the lack of good animal models and the difficulties in obtaining, and working with, human lens material[125]. Consequently, studies trying to explore the molecular causes underlying the vast majority of epidemiological associations for ARC risk factors has not been particularly fruitful. That said, a plethora of cumulative age-related biochemical and biophysical lens changes that can cause lens opacification has been identified (**Figure 1.4.**). The age-related changes may start as early as the prenatal period and affect both the nucleus and the cortex. The age-related opacification is further increased by oxidative stress due to molecular-, photo- and UV-oxidation[28,126,127]. The oxidation-related changes are proposed to be similar to those affecting the other organs in the body. As opposed to other organs such as the brain, age-related changes in the lens are easier to see and measure. Thus cataract is proposed as a marker of oxidative ageing. This role is further backed up by the association of cataract with other markers of ageing and oxidation (e.g. telomere length, systemic inflammation)[128,129].

Osmotic homeostasis and microcirculation: The lens is avascular but its microcirculation system is very important for the function of the lens because it generates a flux of ions and water that circulates through the lens[130]. In ageing, or as result of dehydration, the amount of water entering the lens nucleus via the epithelium and cortex is reduced[131-134]. In the lens cortex, however, an age-related increase in sorbitol and fructose can increase the osmotic pressure, leading to the uptake of water and lens edema[130,135,136]. Animal studies also suggest a link between lens opacities and changes of extracellular fluid volume[137]. The age-related degeneration of the vitreous humor, may result in molecule gradient disturbances, including an increase in oxygen concentration close-to and in the lens[138]. Furthermore, cataract is inevitable consequence of vitrectomy. Hypertension-related inflammation and endothelial dysfunction may alter osmotic homeostasis in the lens[128]. Finally, Mendelian mutations in gap-junction proteins (GJA3, GJA8, etc.), which disturb lens structure and thus microcirculation, cause congenital cataracts[60,64]; but more subtle changes (e.g. variation in expression) may be involved in ARC.

Glucose Metabolism: The activity of many enzymes involved in glucose metabolism decreases with age while the concentration of advanced glycation end products (AGEs) increases, which results in increasing glycosylation of crystallins, the

formation of high-molecular weight (HMW) aggregates and clouding[43]. Furthermore, apart of the effects of UV absorption, the brown pigmentation in advanced cataract may be partially due to a Maillard reaction where a sugar molecule binds to lysin and increases protein cross-linking[139]. Changes in glucose metabolism due to obesity, T2D or other components of the metabolic syndrome may therefore lead to the development of diabetic cataracts, which, in the majority of cases, are not morphologically distinguishable from primary ARC but may differ molecularly[140,141].

Figure 1.4: Ageing-related changes within the lens



This figure, adopted from Harding *et al.*, 2002[54], summarises the age-related changes known to occur in the lens. Glc –glucose; G6P – glucose 6-phosphate; MDA – malondialdehyde; SH – disulphide anions.

Proteins and changes related to protein metabolism: Due to the lack of organelles, the protein and cell membrane replacement and repair mechanisms are limited or lacking altogether, and the lens is dependent on molecular transport from the aqueous humor and lens epithelium[43]. Also with age there is a shift from protein degradation through the proteasome system of degraded to degradation by proteases[43,142,143]. As the lens ages, there is an increase in water–insoluble proteins, which is associated with high levels of yellow chromophores[28]. The chromophores give the lens a yellow coloration, which leads to shift in colour perception and increase in

the formation of oxygen radicals[43,47]. During the lens lifespan, crystallins are truncated through proteolysis and modified molecularly, resulting in protein unfolding, destabilisation and conversion from L- to D-isomers[28,144]. The most common molecular modifications are: thiolation, deamidation, glycation, carbamylation, cys-methylation, phosphorylation, acetylation, intermolecular S-S cross-linking, and aldehyde-amino acid reactions[28,144]. In addition, similar processes affect the cytoskeletal and membrane proteins an lead to their loss[43]. With ageing the oxidative damage of proteins increases[28,126,127], and the levels of enzymes involved in the oxidative stress response and in ubiquitin conjugation decrease[43,142,145]. These mechanisms result in the formation of HMW insoluble aggregates, the levels (>10 times increase between the ages of 20 and 60) and complexity of which increase with age[146]. An additional source of protein oxidation may come from alcohol consumption. Alcohol metabolism via cytochrome CYP2E1 has been hypothesized to generate free radicals causing aggregation of the lens proteins, resulting in subsequent lens opacities[147,148]

Lipids and changes related to lipid metabolism: During ageing there is an increase of lenticular lipid content (from 4.1 to 14.5 mg), mainly due to an increase of oxidation-prone sphingolipids and gangliosides[149]. There is also less cholesterol and phospholipids (0.6-0.8 mg to 0.5 mg) in the deeper layers of the nucleus and the cortex[51,52,150]. The changes in lipids, together with the loss of membrane proteins, results in loss of membrane potential and an increase of Na^{+} and Ca^{2+} ions, at about approximately 50 years of age — the time when cataract formation starts[151]. Alcohol may augment those processes, and in addition, disrupt calcium homeostasis[152,153]. Furthermore, a more systemic lipid dysregulation may also play a role. Obesity-related increases in leptin and high triglycerides and decreased HDL have been implicated in increasing oxidative stress which may contribute to cataract formation[154-156]. The age-related changes that affect lipid composition can be used as an oxidative stress marker of ageing beyond the eye[51].

Changes in antioxidants: A stable redox environment in the lens is achieved by additional scavenger molecules and repair systems located in cell membranes (e.g. vitamin E) and the cytosol (reduced glutathione (GSH), ascorbic acid, lutein and zeaxanthin, cysteine, methionine, glutathione reductase, glutathione peroxidase, thioredoxin and thioltransferase)[157]. Estrogens, which have been implicated in causing the sex-specific differences in cataract formation, have anti-oxidative properties, too[143]. Similar to other parts of the body, the concentration of all these antioxidants drops with age which leads to increased oxidative stress and protein damage[158]. The effect of smoking on cataract, apart from increasing oxidative stress and the accumulation of heavy metal ions[159-161], may also be related to depletion of antioxidants[126].

The drop with age in the concentration of GSH, the most important antioxidant molecule of the lens, is especially pronounced and once cataract develops the rate of GSH decrease is faster[28]. The ageing lens nucleus loses its ability to produce GSH; the synthesis of GSH in the cortex and lens epithelium slows down as does the GSH transport to the nucleus[28]. The role of dietary antioxidants is very important for maintaining a stable redox environment, partly through interaction with GSH. In the absence of GSH, ascorbic acid, which is the second most abundant antioxidant in the lens, can cause oxidative lens damage through its first oxidation product (dehydroascorbic acid)[28]. Decrease in vitamin C concentration, on the other hand, causes an increase in angiotensin-converting enzymes and impacts the proliferation, migration and epithelial-mesenchymal-transition of lens epithelial cells, thus causing PSC cataract[158,162]. Vitamin E, apart from scavenging galactose- and aminotriazole-indused radicals, is also important for GSH homeostasis as it is involved in GSH recycling[163].

All the animal models and in vitro experiments described in this section suggest that disruption of many pathways can all lead to the same end point, that of cataract. It is therefore highly likely that numerous environmental and genetic factors will add up to the complex trait that is ARC.

1.6. Genetics of age-related cataract

Age-related cataract shows familial aggregation and heritability studies for all cataract types have been performed[164-166]. In the TwinsUK cohort, additive genetic factors explained 48% [95%CI: 42%-54%] of nuclear cataract variance, with the remaining variance due to environmental factors and age[167]. Sib-pair analysis in the SEEP found nuclear cataract heritability of 55%, however after adjustment for confounders the heritability fell to 36% [95%CI: 21.0%-50.3%][168]. Results of commingling analysis in BDES, suggested that a single gene of major effect could account for 35% of nuclear cataract[164]. However, such a gene has not been found, and the majority of age-related eye diseases show polygenic inheritance which raises significant doubt about these results. In the TwinsUK cohort, cortical cataract was similarly heritable compared to nuclear cataract with heritability of clinically graded cataract being 53% [95%CI: 45%-60%] and that of digitally graded cataract being 58% [95%CI: 51%-64%][169]. Moreover, the best fitting model suggested a major non-additive genetic contribution (ie. dominant genetic effects)[169]. In the SEEP, the adjusted heritability of cortical cataract was again found to be lower (24%)[170]. The differences in

estimate between the twin- and family-studies are due to the fact the twin studies are known to overestimate heritability[171]. Due to the rarity of PSC cataract in population-based studies, accurate heritability analysis for this trait is difficult.

As opposed to congenital cataract, where over 39 genetic loci and 26 genes have been described, very little is known about the genetics of ARC[172]. Two genes, *GALK1* and *FTL*, have been reported in connection to familial forms of adult-onset cataract. Homozygous mutations in *GALK1*, a gene encoding for an enzyme important for galactose metabolism, cause a mild phenotype characterised by development of cataract postnatally[173,174]. The cataract arises because of osmotic swelling of the lens due to galactitol accumulation. Interestingly, in Asians, a low frequency variant in *GALK1* was associated with ARC[175]. Both dominant and recessive mutations in *FTL* cause iron storage defects which lead to cataract but the penetrance of the different mutations is so varied that cataract may not develop until the 5th decade of life[176,177]. There have also been plenty of predominantly underpowered candidate gene studies that looked at single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) in genes linked to: oxidative stress response (e.g. *GSTM1*, *OGG1*, *XDP*, *XRCC1*, *PARP1*); congenital cataracts (*CRYAA*, *MIP*, *GJA3*, *HSF4*, *SLC16A12*); signalling pathway (e.g. estrogens); structural or cytoskeleton lens proteins[64,178-205]. These studies mostly did not distinguish between cataract types and the associations they identified have generally not been replicated and often reflect publication bias and questionable methodologies.

The first gene robustly associated with a common form of ARC, cortical cataract in particular, was *EPHA2* and this gene has been linked to both congenital and late-onset cataracts[122,123,206,207]. The *EPHA2* locus was initially described using sib-pair linkage analysis and, subsequently, both rare and common coding variants in *EPHA2* were associated with age-related cortical cataract in European and in Asian candidate gene/variants studies[26,122,208-210]. *EPHA2* encodes for a tyrosine kinase receptor, important in development and apoptosis, including UV radiation-induced apoptosis[211]. EPHA2 binds ephrin-A family ligands leading to contact-dependent bidirectional cell-cell signalling, and participates in the cross-talk between RAS-PI3K-ACT and RAS-MAPK signalling pathways, upstream of ERK1/2[122,212], thus promoting cell differentiation.

More recently, a meta-analysis of genome-wide association studies (GWAS) in Asian cohorts found two loci (*CRYAA*, *KCNAB1*) to be associated with age-related nuclear cataract[27]. *CRYAA* encodes for the most abundant structural protein of the lens and mutations in this gene cause congenital cataracts[57,213,214]. *KCNAB1* encodes voltage-gated potassium channel, not previously linked to cataract[215]. Finally, two

studies explored the association between any cataract and SNPs at exome-wide and genome-wide level. A Korean exome-wide study of small sample size (N=249) found only one variant to be suggestively associated with cataract (rs7136356 within *B3GNT4*; $P=6.5x10^{-5}$) and 13/20 most strongly associated SNP to be within glycosylation-associated genes[216]. A case-control GWAS in 7,397 individuals from the eMERGE network followed by age-of-onset analysis in cases only, found 45 variants to be suggestively associated with cataract[217]. Neither study used replication nor made statements on age correction for their analyses. In the latter cohort, gene-by-gene and gene-by-environment analyses were also attempted with doubtful significance[218,219].

1.7. Epidemiology and risk factors

The prevalence of ARCs is dependent on age, sex, and ethnicity. Socioeconomic status affects access to treatment but its effect on cataract formation is uncertain[220]. Comparing the prevalence estimates from different studies is further complicated by the variety of grading systems used[42] and differences in study design and reporting: different age ranges; reporting "any" cataract instead of by type; separating pure and mixed cataracts; combining estimates with those for cataract surgery. Nevertheless there are certain trends, as discussed below, emerging from published studies[166,221-230]. As opposed to prevalence, the data on incidence and progression of ARC are more limited but the incidence patterns resemble those for prevalence[231-238]. Finally, progression of cataracts seems not to be dependent on age or sex but some ethnic differences have been observed[231-233]. These differences, however, may reflect the different grading systems used rather than true ethnic differences.

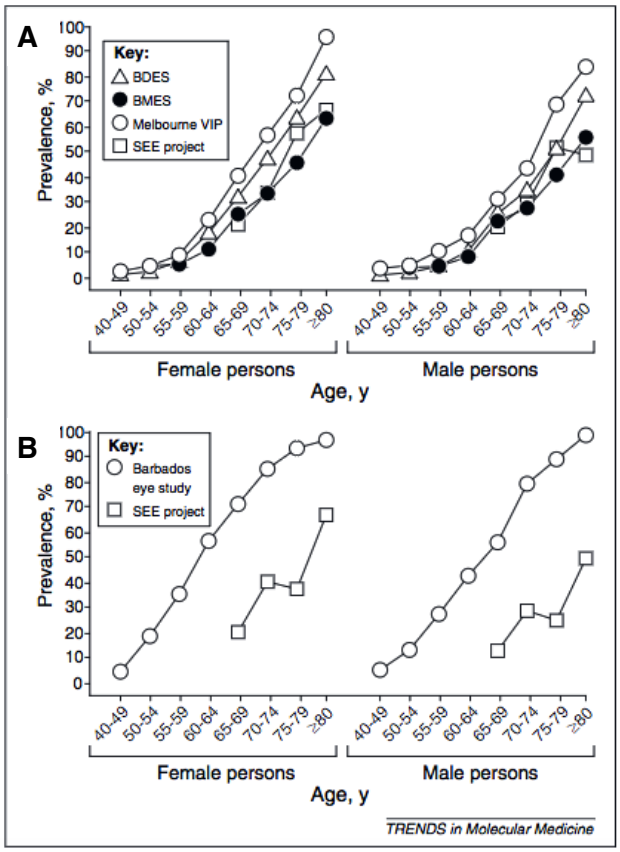
1.7.1. Differences in prevalence between cataract types.

Nuclear cataract is the most prevalent type of cataract*, followed by cortical cataract, while PSC cataracts are comparatively rare. To illustrate the variation, in the Beaver Dam Eeye Study (BDES) overall 17.3% of people had nuclear cataract and 16.3% had cortical cataract[221]. In the Blue Mountains eye study (BMES) slightly older population on average, 53.3% of women and 49.7% of men had nuclear cataract and 25.9% of women and 21.1% had cortical cataract[223]. In India, up to 48% had nuclear

* except for individuals of African decent where cortical cataract is more common than nuclear cataract

cataract and up to 21% had cortical cataract, while the Beijing Eye Study (BES) found the prevalence of any nuclear lens opacity to be 82.0% and that of any cortical lens opacity to be10.3%[227,239]. Some of these differences may reflect different grading systems.

Figure 1.5: Increase in prevalence and incidence of cataract with age



This graph, adopted from Moreau at al. 2012[240], illustrates the increase in prevalence (in %) of cataract as function of age (in years) for both sexes in for people of European (A) and African (B) ancestry (B). BDES – Beaver Dam Eye Study; BMES – Blue Mountains Eye Study; VIP – Visual Impairment Project; SEE – Salisbury Eye Evaluation.

1.7.2. Age and ethnicity

As with other age-related conditions, increasing age is by far the strongest known risk factor for cataract and both the prevalence (**Figure 1.5.**) and the incidence of cataract, including by type, increase with age[166,221-232].

Partially because of differences in healthcare provision, cataracts tend to be more prevalent in individuals of non-European ancestry, and the age of onset can be slightly

lower than 50 years of age in Asians, Africans and Hispanics[222,224,227,241,242]. The reason for this is not clear but, given that genetic risk factors for age-related disease are usually largely shared across ethnic groups, greater lifetime exposure to environmental risk factors (eg. UV light, low antioxidant diets, biomass cooking fuels) is a likely contributor.

1.7.3. Sex

Irrespective of ethnic origin, cortical cataracts are more common in women than in men, while PSC cataract affects equally both sexes. The effect of sex on nuclear cataract is less clear. Several population based studies found that, after adjusting for age and various other confounders, female sex was associated with increased prevalence and incidence of both nuclear and cortical cataracts[166,221,243]. In contrast two other population based studies, the BMES and Barbados Eye Study (BAES), reported cortical cataracts only to have higher adjusted female prevalence[223,244]. Long-reproductive life span (early menarche and/or late menopause) and the use of hormone replacement therapy (HRT) seem to have protective effect against cataractogenesis but the evidence is conflicting[245-249]. For example, in the BDES later age of menopause and longer HRT use were associated with decreased risk, while the Salisbury Eye Evaluation Project (SEEP) found a protective effect for nuclear cataract but not for cortical cataract and only in subgroup analysis[245,247]. The Nurses Health Study (NurHS), however, failed to find any association between current HRT and the presence of any type of lens opacity[247]. In the BMES, HRT use was not associated with 5-year or 10-year incidence of any of the types of cataract but the use of contraceptives protected against cortical cataract (5-year incidence)[248,249]. Finally, a meta-analysis of these and other studies concluded that HRT use decreased the risk of developing any type of cataract for both population-based and case-control studies; and that current HRT use protected against nuclear cataract only[250]. In Asian women, inconsistencies similar to those in European women exist, with some studies finding a protective effect while others reporting increased risk[251,252].

The sex specific differences in cortical cataract risk mimic those in longevity and in other age-related traits (hypertension, cardiovascular (CVD) and neurodegenerative diseases, etc.)[253-255]. Sex-specific differences have been proposed to be biological in nature and driven by sex hormone levels[143]. In particular, low androgen, high estrogen and low testosterone levels seem to have protective effect in various conditions[143]. Furthermore, as exemplified by CVD, after menopause the risk is equal in both sexes or, as is the case with hypertension, women are at a higher risk than men[256]. These

findings do not imply that hormone imbalance does not play a role in disease aetiology in men but rather that, apart from baseline differences in concentration, there might be a difference in threshold effects between the sexes. It is known that prostate cancer patients treated with androgen inhibitors, have elevated incidences of diabetes, CVD and, interestingly, cataract [257,258]. Regarding molecular mechanisms, it is unclear how exactly the sex hormones exert their effects in the lens but the receptors for all of them are present in both the animal and the human lens[259,260]. One possibility is through anti-oxidant effects (estrogen) or through the suppression of antioxidant defences (androgens, testosterone)[261,262]. Estradiol, the predominant form of estrogen, has been shown to protect cultured human lens epithelial cells (LEC) against oxidative stress[263,264]. Another possible mechanism in LEC relates to mitochondrial health. It has been demonstrated that estradiol causes rapid activation of extracellular signal-regulated kinase (ERK1/2), a member of the mitogen-activated protein kinase (MAPK) survival signalling pathway, and that ERK2 silencing causes a dramatically increased membrane depolarisation of mitochondria [265,266]. In mice, estrogen has also been shown to protect lenses against TGF-β induced cataract[267,268]. Finally, similarly to the results of epidemiological studies, the use of HRT in animal models of cataract and of other disease showing sex-specific effects has yielded conflicting results[268,269].

1.7.4. Tobacco smoking

Tobacco smoke is a complex chemical mixture which can adversely affect human health through several mechanisms, discussed below. Tobacco smoking (smoking status, pack-years) has been extensively studied as a risk factor for ARC and some of the studies that explored the relationship between smoking and cataract will be discussed here.

The observation that smoking increases cataract risk was first noted in case control studies[270,271]. Later, several studies used case-control design to specifically look at different cataract subtypes. The Lens Opacities Case Control Study (LOCCS) found increased risk for nuclear cataract only, as did the Age-Related Eye Disease Study (AREDS)[272,273]. A Greek hospital-based study found association between smoking and both nuclear and PSC cataracts[274]. Further evidence was provided by prospective and cross-sectional population based studies. For example, The Reykjavik Eye Study (RES), a cohort study of comparable size, found association between smoking and nuclear cataract only while The Physicians Health Study (PHS) found increased 5-year incidence of nuclear and PSC cataracts in heavy smokers[275,276]. Finally, a recent meta-analysis concluded that smoking statistically significantly increases the risk of nuclear

cataract (OR_{cohort} [95%CI] = 1.7 [1.5-1.9], $OR_{case\text{-}control}$ [95%CI] = 1.9 [1.5-2.4]), and that there was a marginally significant association with PSC but no association with cortical cataract[277].

Dose-response and reversibility effects were also demonstrated. For example, the City Eye Study and Framingham Eye Study, found that heavy smokers had increased risk of nuclear cataract compared to light smokers[278,279]. In PHS, compared to smokers, past smokers had a statistically significant reduction in risk of cataract diagnoses and cataract extraction[280]. The effect, however, was primarily due to lower total cumulative dose, except for cataract surgery were there was a small dose-independent effect[280]. Unlike cataract formation, the effects of smoking on cataract progression are not well understood. Only one study of modest size, the Chesapeake Bay Eye Study (CHBES), looked at progression and suggest that odds of progression increased for smokers versus ex- and non-smokers; and 18% increase in risk for each pack year of smoking[281].

Tobacco smoking also increases the risk of cataract formation in non-European populations, although this is predominantly based on results in Asian populations, as studies in African and Latin-American populations are scarce[282,283]. Similarly to the majority of studies in European ancestry populations, in individuals of Chinese ancestry smoking increases the risk of nuclear cataract formation only[238,243,284]. Studies in ethnic Indians[285,286] and Malaysians[287], found similar effects for nuclear cataract but, unlike the previous studies, they also found association between cortical cataract and smoking. For example, an Indian cross-sectional population based study, found that smoking increased the odds of nuclear and cortical cataracts, but not of PSC cataract[285], while cross-sectional association between smoking and all the cataract types was reported by the Singapore Malay Eye Study (SiMES)[287].

The molecular mechanisms behind the association between nuclear cataract and smoking are not extensively studied but it is assumed that the smoking-related increase in oxidative stress is responsible for the increase in nuclear cataract risk. It is well established, both in vitro and in vivo, that cigarette smoke induces reactive oxygen species (ROS), which in turn cause oxidative damage to lipids, proteins, and DNA[288]. In many cases, the initially generated reactive intermediates convert cellular constituents into second-generation reactive intermediates (e.g., acrolein, 4-hydroxynonenal) capable of inducing further cytotoxic and genotoxic damage, resulting in chain reaction of free radicals[288]. ROS are capable of activating numerous redox sensitive signaling pathways that modulate cellular responses, such as inflammation, which may itself result in the formation of endogenous oxidative species[288]. Further to the oxidative properties of cigarette smoke, smoking-related increase in the concentration of heavy

metals in the lens, has also been proposed to play role in cataractogenesis[160,289-292]. Those heavy metals are also prosed to increase oxidative stress damage by catalysing Fenton reactions[290].

Another possible mechanism for the effect of smoking on cataract is through epigenetic changes affecting crystallin expression or the expression of genes with antioxidant roles[293-295]. The field of lens epigenetics is relatively new but the ways in which smoking cases epigenetics changes have been extensively studied in other conditions. Smoking has been consistently shown to trigger genome-wide epigenetic changes in methylation levels[296,297]. For example gene enrichment analysis have shown smoking related epigenetic changes (either hyper- of hypomethylation of CpG cites) to be more common in genes associated with coronary heart disease risk factors (hypertension, lipid levels), pulmonary traits and cancer[298]. Another example is multiple sclerosis where smoking-associated hypomethylation has been detected at the aryl hydrocarbon receptor gene CpG[299]. The aryl hydrocarbon pathway is known to exert environmental control on the immune response and, in particular, on neuronal inflammation [299].

1.7.5. UV radiation

Studies looking at a link between exposure to UV radiation, especially UV-B, and cataract formation had varying degrees of success. UV radiation has repeatedly been reported as a risk factor for cortical cataract formation but the picture for the other types of cataract is less clear[300-304], and defining exposure or disentangling the roles of the different light wavelengths is tricky.

The first study to look at the relationship between UV radiation exposure and cataract in two population-based surveys (the Model Reporting Area for Blindness Statistics and the National Health and Nutrition Examination Survey 1971–1972) noted that the ratio of cataract cases to controls was significantly higher in areas with large amounts of sunlight for people aged 65 and older[305]. A later study, using the Medicare claims data, found that the odds of cataract surgery were also higher in southern than in the northern states[306]. Moreover, The Alienor Study found subjects in the upper quartile of lifetime ambient UVR exposure to be at increased risk for cataract extraction[307].

Contrary to those findings, case-control studies have mostly failed to find an association[272,308-311] but this can be due to small sample-size, the ways used to calculate exposure, and selection bias. An exception is the RES where more than 4 hours outdoor daily exposure before the age of 50 was associated with increased risk

of moderate and severe cortical cataract[312]. Interestingly, a cross-sectional population-based study of Finnish individuals approximately at the same latitude as Reykjavik, albeit much older (≥70 years), did not find an association[313]. Studies from Australia showed that the risk of cataract increases with higher UV radiation dose in populations of both Aboriginal[314,315] and European descent[316]. These results have been supported by several other studies[316-322].

As UV radiation increases with increase in altitude, a positive association between altitude and cataract formation can be expected. Indeed, several studies have found increased cataract prevalence at higher altitudes[323-327] but there have also been some conflicting results. A study in Punjabis found less cataract in people living in the regions at higher than at lower altitude, and a study in Nepal found that cataract prevalence was negatively correlated with altitude but positively correlated with sunlight exposure[328,329]. The first study was of moderate size but the difference in altitude between the sampling locations was small, and cloud cover, which seems to have a protective effect[330], may have biased the results. The discrepancy in the second one is more interesting but may point to population specific high-altitude adaptation differences[331,332].

The cataractogenic effects of UV-light have been studied in animal models and those studies suggest that UV light can cause lens damage in various ways. UV light (both UVA and UVB) has oxidative properties[333]. Lenses of mice who were exposed to UV light showed increase in oxidative stress markers[334]. GSH level and glyceraldehyde 3-phosphate dehydrogenase activity decreased while mixed disulfides concentration, thioltransferase and thioredoxin activity increased[334]. UV-B irradiation caused changes in lens proteins such as side chain oxidation and polypeptide cleavage of γD-crystallin and βL-crystallin; and formation of cross-linked βL-crystalline aggregates[335]. These changes led to the occurrence of amorphous crystalline aggregates and of amyloid fibres[335]. Furthermore, AGEs are proposed to be important contributors during cataract development due to their cross-linking properties[333]. UVA radiation excites kynurenine to photo-oxidised ascorbate and reduces ascorbate which in turn can forms AGEs[336]. UVA seems also to increase apoptosis in animal models. Exposure of high doses of UV-light induced apotheoses in rodent LEC and the LEC's phagocytic capabilities were lost[337,338]. The apoptosis caused intracellular calcium ion ($Ca2^{+}$) elevation, inhibited $Ca2^{+}$-ATPase activity and decreased the expression of PMCA1, a plasma membrane calcium, with a role in calcium homeostasis[339]. This suggests that PMCA1 may play important roles in UVB-induced cell apoptosis[339]. Finally, UVB exposure in mice was associated with intraocular inflammation and increase in concentration of circulating Interleukin-1b and interleukin-6[340].

1.7.6. Dietary factors and alcohol consumption

The majority of epidemiological studies, discussed below, explored the relationship between cataract and diet focusing on specific dietary factors (e.g. vitamin C) and data on diet types or intake of food groups is less studied. As opposed to macronutrients and vitamins, the role of dietary minerals in cataract formation is poorly understood[341-345], with only iron, calcium, zinc and coper proposed to have some effect[272,330,341,343,346].

Regarding alcohol consumption, the different studies show inconsistent direction of effect per cataract type. For example, in the NurHS, the odds of nuclear opacity increased by 30% for each 10g increase in total alcohol intake, after adjustment for age and vitamin C intake, but there was no statistically significant relationship between alcohol intake and cortical cataract[153]. Conversely, in the BMES cohort moderate alcohol consumption was associated with reduced prevalence of cortical cataract but there was no association with nuclear cataract[347]. There results of the different studies, however, are not always consistent and several studies report no association for any of the cataract types[166,225,348,349]. Similarly, studies that focused on cataract extraction can be divided into those that find alcohol is a risk factor[350] and those that either find no association[341,351,352] or a protective effect[353]. Furthermore, a recent meta-analyses failed to show any association between alcohol consumption and ARC[354,355]. Similarly to the evidence from the epidemiological studies, there is lack of molecular evidence for a causative role of ethanol on cataract formation. It seems that, while alcohol-derived sugars play a role in certain types of cataracts (eg. galactosemic[356] or diabetic cataracts[140]), they probably do not in the common forms of ARC.

Macronutrients: The results of several small studies have proposed high carbohydrate intake to be associated with an increased risk of developing cortical cataract[357-359]. However, several larger studies found no cross-sectional or longitudinal association between carbohydrate intake and cataract[343,360,361], while a small Greek hospital case-control study even found carbohydrates to protect against any cataract[362]. Finally, a recent meta-analysis study, found only a marginal association (OR [95%CI] of 1.4 [0.99-1.9]) between carbohydrate intake and cortical cataract[363]. It has been suggested that an association between carbohydrates and cataract could be driven by fibre intake, however fibre seems not to influence the risk of cataract, except possibly of PSC[343,361,364].

Another way to approach the problem of whether carbohydrates influence

cataract pathogenesis is to use the glycemic load as a predictor. The glycemic load is the weighted average of the glycemic indices of individual foods, multiplied by the percentage of energy as carbohydrate[365]. The glycemic index is the incremental area under the glucose response curve after a standard amount of carbohydrate from a test food relative to that of a control food (either white bread or glucose) is consumed[366]. In AREDS, dietary glycemic index was associated with a higher prevalence of all pure nuclear opacities[358]. The BMES found a nominally significant association between the glycemic index and 10-year incidence of cortical cataract but not between the glycemic load and the incidence of other cataracts[360]. Several other studies either failed to find effect of glycemic load/index on cataract or found marginal effects, without dose-response associations[357,367]. An Italian case-control study found an association between glycemic load and cataract extraction, while, a much bigger study using combined data from the NurHS and the Health Professionals Follow-Up Study did not find an association between glycemic load and cataract extraction[368,369]. Wu *et al.*, after performing meta-analysis, concluded that there was association between higher glycemic index and nuclear cataract risk (OR [95%CI] of 1.23 [1.03-1.46]) and a significant dose-response relationship between carbohydrate intake and the risk of cortical cataract[363].

Very few studies looked at the association between protein intake and cataract. A retrospective analysis in the BMES, indicated that subjects who consumed a median of 99.2g of protein per day had a 50% reduction in risk of prevalent nuclear cataract compared with those in the first quantile (62.3g) of protein intake[343]. In the same population, high protein consumption was associated with PSC incidence only[361]. A similar cross-sectional effect was found in Chinese subjects[359]. In EPIC, however, high intake of protein was associated with increased risk of cataract formation[370]. High protein intake was also associated with higher risk of cataract in an Indian study[328].

Finally, although various studies explored the relationship between fat intake and cataract, relatively few associations have between reported. The EPIC study reported increased risk of any cataract with elevated blood levels of cholesterol[364], while the prospective SUN cohort had decreased risk for any cataract in people who consumed high levels of polyunsaturated fatty acids (PUFAs)[371]. The BMES found cross-sectional protection of PUFAs for cortical cataract[343] and, also that omega-3 PUFAs were protective against nuclear cataract incidence[361]. Conversely, in the NurHS (Boston Area) high total intake of PUFA resulted in 2.3-fold increased risk of nuclear cataract but this could have been driven by a strong negative association between linoleic acid intake and nuclear cataract[372,373].

Vitamins and carotenoids: The role micronutrients play in cataract formation

has also been studied. Given that, as discussed later, oxidative stress is proposed to play an important role in cataract pathogenesis, studies have mostly focused of the effects of dietary antioxidants (vitamins C, E, and A, carotenoids) and vitamins in general.

Three strategies have been deployed when studying the effect of vitamin C on cataract formation. These are measuring: dietary vitamin C intake[272,374-378]; serum vitamin C concentrations[375,376,378-380], and vitamin C supplementation[375,377,381]. Randomised clinical trials of supplement have also been performed[382-384]. These studies sometimes have conflicting results, as discussed below, but in most studies, vitamin C intake does seem to protect against nuclear cataract formation. Furthermore, vitamin C has also been found protective against cataract progression[385,386]. The LOCCS looked at all three types of cataract but found vitamin C intake to be protective against nuclear cataract formation only[272]. Similar results were obtained by a Spanish case-control study[376], while an American-Italian clinical trial found protective effect at baseline for both nuclear and PSC cataracts[380]. Inverse association between vitamin C intake and cataract formation have also been found in population based studies. The NurHS found that intake of vitamin C was associated with lower prevalence of nuclear opacities only and a protective effect of long term (>10 years) intake of vitamin C supplements[375]. The BMES explored the association between incidence of cataract and vitamin intake, and found that vitamin C intake or combined antioxidants (vitamins C and E, β-carotene, and zinc) reduced cataract incidence[377]. However, other prospective cohort studies and randomised clinical trials have found no effect overall[374,379,382-384,387] or protective effects only in subgroups[374,387]. Somewhat surprisingly, given the rest of the literature, two Danish population based studies found long term vitamin C supplementation to increase cataract risk[388,389]. With the exception of two studies[376,378], vitamin C intake seems not to be associated with cortical cataract. Finally, a meta-analysis of 20 studies for vitamin C intake and 10 studies for serum ascorbate, concluded that vitamin C intake reduced the risk for any cataract (RR [95%CI] = 0.81 [0.71-0.94]), as well as for nuclear and PSC cataracts[390]. The association between vitamin C intake and cataract was primarily driven by the case-control studies, and the studies in Americans and Asians. An inverse relationship between cataract and serum ascorbate was found irrespective of study design or geographic location (RR [95%CI] of 0.70 [0.56–0.88])[390].

Ascorbate (vitamin C) is found at high levels in the lens where it is capable of quenching oxygen and, in addition, acts as a strong reductant and scavenger of ROS[391-393]. Ascorbate has been shown to decrease membrane damage in diabetic rats and photoperoxidation of lens membranes[394,395]. Furthermore, ascorbic acid prevents

light-induced damage to cation pumps in the lens; plays a role in lens development and maintenance of transparency during development; and has been shown to protect the rat lens against selenite-induced cataract[396]. Lenticular levels of ascorbate declined with age and the ratio of ascorbate to dehydroascorbate decreased in cataractous lenses compared to clear lenses[396]. Finally, aside from its reducing properties, ascorbic acid is also essential for collagen synthesis and has anti-inflammatory properties[397].

Vitamin E encompasses a family of molecules known as tocopherols, of which α-tocopherol (highly concentrated in olives and sunflower seeds) is the most biologically active. It is also the major tocopherol in the European diet, whereas γ-tocopherol (soy, corn) is the major tocopherol in the American diet[163]. Compared to both vitamin C and A, the role of vitamin E in cataractogenesis is more uncertain and the majority of studies did not find an association between dietary vitamin E and cataract or, when they did, it was only in univarable analyses[343,374,376,398,399]. A recent meta-analysis of 27 studies, however, concluded that vitamin E intake was protective against cataract formation (OR [95%CI] = 0.73 [0.58-0.92]), as were supplementation and serum levels[400]. The pooled odds in this analysis was calculated using the odd ratios for different types of cataracts and any cataract together which may not be optimal. Vitamin E supplementation and vitamin E plasma levels appear to have protective effects[375,379,381,385,386]. In the NurHS, high levels of vitamin E in plasma and duration of vitamin E supplementation were inversely associated with reduced risk of nuclear cataract cross-sectionally[375]. Moreover, in the same cohort duration of vitamin E supplementation decreased the 5-year incidence of nuclear cataract[386]. Similarly, Robertson *et al.* found supplemental vitamin E to be associated with reduced risk of cataract, while Vitale *et al.* found high levels of plasma tocopherol to be associated with reduced risk of nuclear cataract[379]. Meta-analysis also found tocopherol in plasma to be inversely associated with cataract (OR [95% CI] = 0.75 [0.58-0.96])[401]. Similarly to vitamin C, in randomised clinical trials, vitamin E supplementation alone or in combination with other vitamins failed to find an effect[302,382,383,402,403]. Finally some studies have reported an increased risk of cataract in association with vitamin E intake. In the MVIP and BMES cohorts, vitamin E intake was associated with an increase in PSC cataract risk[166,377]. Furthermore, in BMES, high vitamin E levels were also associated with increased prevalence of cortical cataract, and of any cataract[377].

Vitamin E functions as a lipid antioxidant and a free radical scavenger, and may help to maintain GSH levels and to reduce protein–protein linkages in the lens[404,405]. There is evidence suggesting that vitamin C recycles vitamin E from α-tocopheryl thus pointing to synergetic relationship between the two vitamins[396]. The delaying or preventive effect of vitamin E on cataractogenesis has been shown in galactose

cataract mouse models but, just as with the epidemiological studies, vitamin E supplementation in mice models of idiopathic ARC showed mixed results[396]. Moreover, a recent study in humans has suggested that there is no correlation between the concentration of vitamin E in lens and blood (plasma or red blood cells) and, thus, that plasma α-tocopherol may not be a clinically relevant marker for cataract risk[406].

As discussed below, the role of other vitamins in protecting against cataract formation has been addressed trough epidemiological studies but, as opposed to vitamin C and E, the molecular effects of all those vitamins on cataract formation are very poorly understood.

Vitamin A is a group of unsaturated organic compounds that includes retinol, retinal, retinoic acid, and several provitamin A carotenoids (α-carotene, β-carotene, and β-cryptoxanthin). The majority of studies that explored the effect of vitamin A have shown an inverse association with cataract, except for a study in the EPIC cohort where high intake of vitamin A was associated with increased risk of "any" cataract[364]. As with vitamins C and E, supplement trials of vitamin A have largely failed to find an effect[382]. To illustrate this, the BMES showed dietary vitamin A to protect against nuclear cataract formation[343], while in the BDES that was true only for men[407]. In the same cohort, participants who supplemented with vitamin A had a reduced risk of cortical cataract[346]. In the POLA study, higher plasma retinol level was associated with decreased relative risk for nuclear cataract, mixed cataract, and cataract extraction[408]. In a cross-sectional study of the North India subset of INDEYE, blood levels of retinol had a protective effect against nuclear cataract only and that was also true for β-carotene[409]. In the South India subset of INDEYE, retinol protected against PSC (and "any" cataract)[378]. Gale *et al.* found that α-carotene decreased the risk of nuclear cataract in British women (aged 66-75 years) living in Sheffield[410]. Finally, meta-analyses on the effect of dietary, supplemental and plasma antioxidants on cataract found retinol*, dietary β-carotene and plasma α-carotene to be inversely associated with cataract[401,411].

The B vitamins are water soluble organic compounds that play important roles in metabolism but only vitamins B1-3, B9 and B12 have been studied in relation to cataract. Cumming et al found 3 of the B vitamins to be inversely associated with nuclear cataract[343]. Similar results but in men only were obtained in the BDES cohort[407]. Jacques *et al.* found the intake of the three B vitamins was inversely associated with nuclear cataract progression; and that dietary B9 (folic acid/folate) was protective against nuclear cataract formation[375,386]. In the AREDS cohort niacin

* dietary, supplemental and plasma

protected against incident (10 years) nuclear cataract[412]. Sen *et al.* found decreased blood levels of folate and vitamin B12 (cyanocobalamin) in cataract patients compared to controls[413]. Moreover, a combination of niacin and riboflavin protected against nuclear cataract formation (44% reduction in prevalence) in a supplementation trial, albeit in a poorly nourished population[414]. Regarding the other two types of cataract, the association between dietary intake of vitamins of the B group and cortical cataract is unclear and there seems to be no association between these vitamins and PSC cataract. Indeed, The LOCCS was the only study to find an inverse association between riboflavin and niacin and, in this case, pure cortical cataract[272]. Finally, in AREDS cohort vitamin B12 was associated with decreased risk of incident cataract[412].

The role of vitamin D in relation to cataract is poorly studied and that of vitamin K not at all. In a secondary analysis of a randomised trial of the Mediterranean diet , high intake of dietary vitamin K1 was associated with a reduced risk of cataract surgery in Spanish men and women (mean age at baseline of 66.3 years)[415]. The Korea National Health and Nutrition Examination Survey suggested that high serum 25-hydroxyvitamin D levels are protective against cataract in men but not in women[345]. In the Carotenoids in Age-Related Eye Disease Study (CAREDS), an ancillary study of the Women's Health Initiative, serum 25-hydroxyvitamin D levels were not associated with nuclear cataract either[416].

Finally, carotenoids are more than 400 organic pigments that are produced predominantly by plants. The major carotenoids in the eye are lutein and zeaxanthin and they likely protect against cataract formation. The majority of studies on dietary[417,418] or plasma[409,410,419,420] carotenes report estimates that, although not statistically significant, point to protective effects overall or in subgroups. The couple of studies that do report statically significant protective effect are discussed here. In the BDES dietary lutein, and in the AREDS and CAREDS studies a combination of dietary lutein and zeaxanthin, were inversely associated with nuclear cataract[374,421]. A Finish population-based study found that individuals with high plasma lutein (> 0.27μmol/L) and zeaxanthin (>0.041μmol/L) were protected against nuclear cataract formation[422]. Similarly, in the CAREDS cohort, plasma lutein plus zeaxanthin were inversely associated with lower nuclear cataract risk[421]. Gale *et al.*, reported an inverse association between plasma lycopene and cortical cataract[410]. In INDEYE, both plasma lutein and zeaxanthin were inversely associated with “any” cataract[378]. Finally, meta-analyses of studies that looked at plasma antioxidants concluded that lutein (OR [95%CI]: 0.75 [0.65-0.87]) and zeaxanthin (OR [95%CI]: 0.70 [0.60-0.82]) but not lycopene (OR [95%CI]: 0.86 [0.65-1.15]) protect against cataract formation; and that there was a dose-response relationship[401,423].

Analogous to the macula, lutein and zeaxanthin are the only carotenoids present in the human lens but at substantially lower concentration in the lens relative to the macula[424]. The molecular mechanisms behind their association with cataract are not well studied but, similarly to the macula, they are hypothesised to serve as high energy blue light filters and as functional antioxidants[424].

In summary, diet seems to have stronger effect on nuclear opacities than on the other types of cataract and, not surprisingly, all dietary effects seem much stronger in malnourished as opposed to well-nourished populations. In addition, the relationship between cataract and metabolic syndrome (discussed later), may partially be mediated by dietary factors. Although studying separate nutrients have been successful, little attention has been given to diets as a whole[364,425-428]. Each nutrient can be confounded by other nutrients and other confounders of a "healthy" diet as a whole, so measures of the overall diet may be a better measure of its effects, and studies in, for example, cardiovascular disease have recently concentrated on dietary indices rather than individual subcomponents[429-431]. Given the few studies to date that looked at entire diets or dietary indexes, I intend to study the role of diet on cataract formation in TwinsUK.

1.7.7. Type 2 Diabetes and metabolic syndrome

Diabetes and cataract occur more often together than by chance, and T2D is associated with increased risk of developing cataract[432,433]. In the NurHS presence of T2D was associated with higher risk of developing PSC, but neither diabetes nor measures of adiposity were related to the prevalence of cortical and nuclear opacities[434]. BDES reported an increased prevalence of PSC in diabetics but there was also an association with cortical cataract[435]. In the same cohort, T2D was associated both with increase in 5-years incidence and with progression of cortical and PSC cataracts[436]. In the MVIP cohort, T2D duration was cross-sectionally associated with cortical and nuclear cataract but not with PSC cataract[166]. Other studies, that explored the association between T2D and cortical[284,437-439], nuclear[438,440,441] and PSC[284] cataract, found similar results. Furthermore, T2D is also found to be a risk factor for cataract formation in individuals of non-European descent; and diabetics with Asian and African descent have a slightly higher risk than non-diabetics[284,437]. Finally, meta-analysis of T2D in cataract, involving 8 population-based studies, concluded that the risk of cortical cataract and PSC cataract (OR [95% CI]: 1.68 [1.47-1.91]) and 1.55 [1.27-1.90] respectively) but not of nuclear cataract (OR [95% CI] = 1.36 [0.97-1.90]) was significantly higher in T2D patients[432]. In addition to T2D diagnoses, impairments

in glucose tolerance[442-444] and fasting glycaemia[440], high glycated haemoglobin[283,443,444] and poor glycaemic control[283,408,439,445,446] may also increase cataract risk.

From a molecular biology prospective, there can be several links between diabetes and cataract, again including oxidative stress. As a consequence of impaired glucose metabolism in diabetes, the activity of aldose reductase (the enzyme responsible for fructose formation from glucose) increases[447]. That then leads to increased activity of the sorbitol pathway and intracellular sorbitol accumulation in the lens; resulting in osmotic stress[447]. Imbalances in glucose metabolism, in the absence of T2D, that lead to osmotic stress may similarly be a driving force in the development of cortical cataract. Accumulation of AGEs is also increased in both diabetic and ARC[43,447]. More importantly, diabetes increases oxidative stress and oxidative stress is the one of the main drivers behind ARC[447]. Thus, recently, the NRF2-Keap1 signal transduction pathway, one of the main cellular defence mechanisms against oxidative stresses, has been proposed as a target for prevention of both age-related and diabetic cataracts[448].

Given the role of T2D in cataract, a question about the role of metabolic syndrome of cataract formation arises. Metabolic syndrome is defined as a cluster of risk factors for cardiovascular diseases (CVD) and type 2 diabetes mellitus (T2D) which occur together more often than by chance alone[449]. These risk factors are: hypertension, dyslipidemia (raised triglycerides and lowered high-density lipoprotein cholesterol), raised fasting glucose, and central obesity[449]. Even though CVD, T2D and their underlying risk factors have been associated with eye diseases[450], the relationship between metabolic syndrome and the ageing eye has only recently been addressed[450-452]. Other lines of evidence connecting metabolism with ocular health are the eye manifestations (e.g. cataract, glaucoma) of inherited metabolic disorders[453].

In population-based cross-sectional studies, metabolic syndrome was associated with increased cataract prevalence[454,455], albeit only in certain subgroups of participants. An increased risk of cataract surgery was observed in a hospital-based case-control study[456] and in a prospective cohort[457]. A population-based incidence study in BMES showed that the presence of metabolic syndrome at baseline was associated with statistically significant increase in 5-year-incidence of cortical cataract, suggestive increase in PSC and no change for nuclear cataract[458]. The same study did not find any increase in 10-year-incidence for any of the cataract subtypes[458]. An earlier larger study in the same population, however, did find an increase in risk of all three cataract subtypes for cortical, nuclear and PSC respectively[440]. The association between metabolic syndrome and cataract is present in Asian populations, too. In the SiMES cohort, cortical cataract was also significantly associated with metabolic syndrome[459]. The study did not find an association with nuclear cataract or PSC

cataract but this could have been partially due to fewer affected individuals in those groups[459]. While a Korean population based study found association between metabolic syndrome and cortical cataract in women but not in men and the inverse was true for nuclear cataract[460]. Finally, a dose–response relationship was also observed between increasing number of metabolic syndrome components and cataract in several studies[456,457,459].

Chapter 2. Aims and objectives of the thesis

Age-related cataract is a complex trait and to prevent or delay its onset, or to develop new therapies for cataract, a better understanding of the genetic and environmental factors and of how these factors interact is required. In addition, it is increasingly clear that pathogenesis of ARC is intertwined with systemic factors related to ageing and holistic approaches taking into account the systemic nature of the cataract risk factors are needed. High-throughput 'omics' technologies have recently become available and affordable, and are enabling detailed studies of the molecular changes and biological pathways involved in ageing processes[461]. Their use in ARC research is long overdue. Utilisation of omics is likely to improve our capability to identify accurately molecular mechanisms leading up to cataract formation in the ageing eye. This current thesis describes the implementation of various Omics platforms to:

Aims: This thesis aims to explore the genetic variants underlying the pathogenesis of the common forms of ARC. In addition the role of diet on cataract formation and progression is also explored.

Hypotheses:

There exist genetic variants associated with age-related nuclear cataract.

There exist genetic variants associated with age-related cortical cataract.

Age-related nuclear and cortical cataracts share genetic risk factors.

Cataract progression is heritable.

Dietary factors are associated with cataract progression.

Healthy diet is inversely associated with at least one ARC type.

There exist metabolite, measured in blood, that are associated with healthy diets and those are inversely associated with cataract severity for at least one ARC type.

There exist gut microbes that are associated with healthy diets and those are inversely associated with cataract severity for at least one ARC type.

The association between healthy diet and ARC is mediated by the effects of metabolites or the gut microbes.

Objectives:

1. Explore whether nuclear and cortical cataract share genetic risk factors
2. Identify common genetic variants associated with ARC by performing genome-wide association studies and the meta-analysis of those studies
3. Identify whether low frequency and rare genetic variants from whole-genome sequencing data are associated with ARC
4. Estimate the heritability of cataract progression.
5. Determine what nutritional factors underlying cataract progression.

6. Identify correlation of dietary scores with cataract severity and explore related metabolome or microbiome signatures
7. Determine if the dietary effects are mediated through circulating metabolites or the gut microbiome

The findings of this thesis will be presented as follows. In **Chapter 3** the cohorts that contributed data to the analysis presented in this thesis will be introduced. The data collection I have performed in two of the cohorts (TwinsUK and the Rotterdam Study) will also be described. In **Chapter 4** I explore whether the two most common types of ARC share genetic risk factors with each other or with other age-related diseases. This is done by using various methods for calculating heritability. Then, the role that common and rare genetic variants play in nuclear and cortical cataract formation is determined by GWAS and burden tests. The common variant analyses were done using two cohorts of European ancestry — TwinsUK and Rotterdam Study, while the rare variant analyses were done using TwinsUK data only. In **Chapter 5** multi-ethnic meta-analysis of nuclear cataract GWAS are presented. In **Chapter 6** the following is explored: 1) the effect of micronutrients intake on cataract progression; 2) whether the effect of diets on cataract formation might be mediated through changes in blood metabolites or the gut microbiome. In the final chapter (**Chapter 7**), the findings reported in each chapter are drawn together and the overall implications of this thesis are discussed.

Chapter 3. Cohorts description. Data collection in the TwinsUK and Rotterdam studies

3.1. Introduction

To aid the understanding and interpretation of the analysis and results presented in the later chapters of this thesis, the cohorts which provided data used in the thesis will be introduced in the present chapter. Here the focus will be on the methods and quality control (QC) steps that were undertaken in order to obtain the various omics datasets that I have used for my analysis. Unless stated otherwise, the data has been provided to me in its final QCed form by my colleagues from the Department of Twin Research and Genetic Epidemiology or from the studies that took part in the International Cataract Genetics Consortium.

In addition to the above, I will describe my efforts in terms of data collection for two of the cohorts — TwinsUK and the Rotterdam Study; and I will present basic summary statistics for the phenotypes I obtained through those efforts.

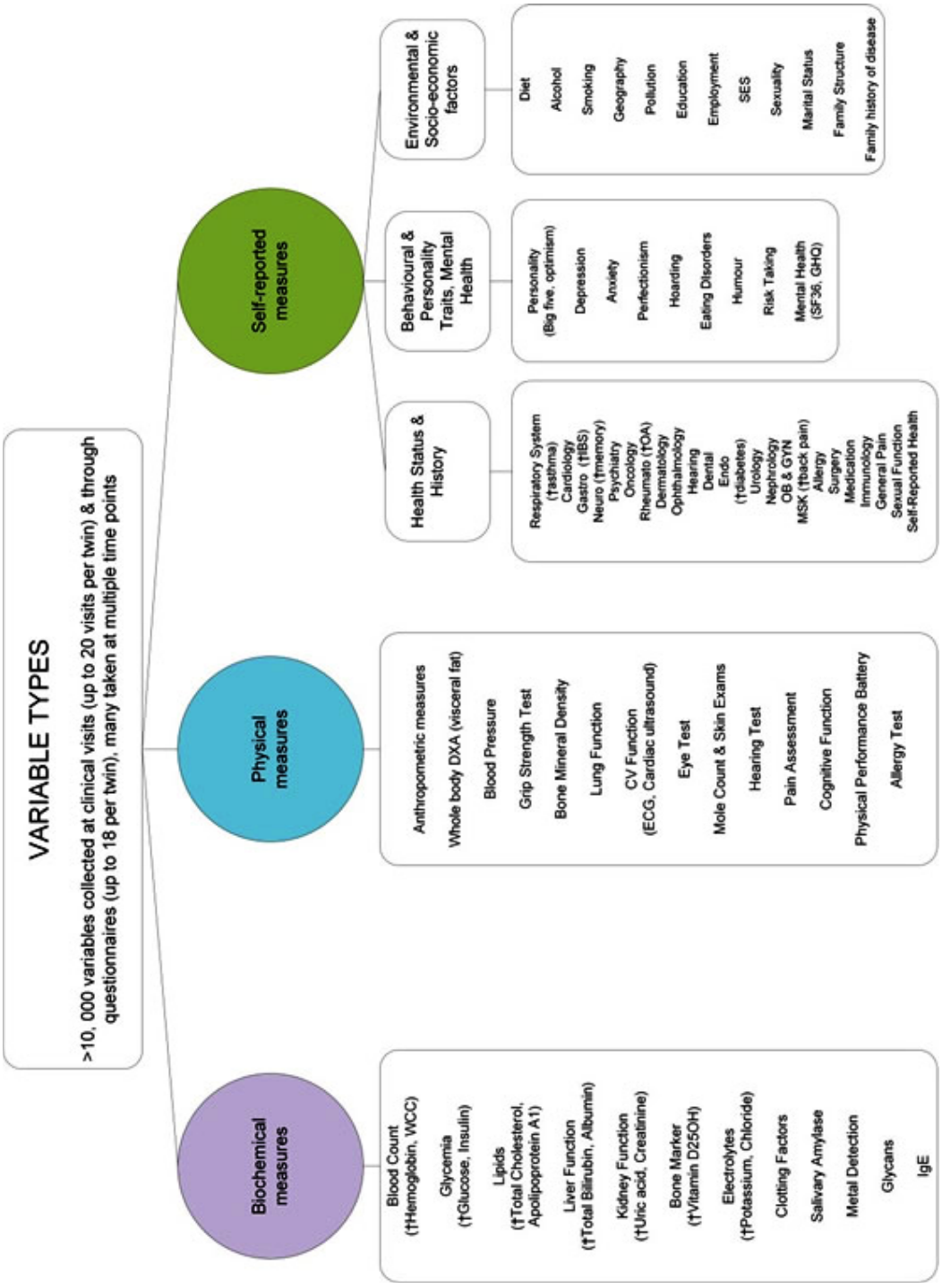
3.2. Twins UK

3.2.1. Cohort description

The TwinsUK cohort is an adult unselected twin registry of over 13,500 British twins (including a small sample of parents and siblings, number of individuals (N)=1,883) recruited from the general population, through national media campaigns in the United Kingdom, and representative to the broader population in terms of disease-related and lifestyle characteristics (http://www.twinsuk.ac.uk)[462-465]. The registry was initiated in 1993 and is still actively recruiting participants. At present over 9,000 twins are actively participating. The twins have a mean age of 55 year (range: 16-100 years) and the ratio of monozygous to dizygous twins is approximately 1:1. The twins are predominantly female (80%) of white British ancestry (80%) for historical reasons: the original aim of the registry was to allow investigation of osteoporosis and osteoarthritis, conditions with higher prevalence in women[463]. Since then, the TwinsUK studies have focused on the genetic basis of complex diseases (cardiovascular, metabolic, musculoskeletal, and ophthalmologic diseases), which has broadened to include the complex healthy ageing process. In addition twins of ethnically diverse backgrounds are now actively recruited. Clinical, physiological, behavioural and lifestyle data is collected at either twin visits to the DTR or via self-administered questionnaires, which volunteers complete either once or twice a year via the post or email (**Figure 3.1.**,

Table 3.1.). Longitudinal data is available for a wide range of the phenotypes. All the

Figure 3.1. Phenotypes available in the TwinsUK cohort



This figure* summarises the different phenotypes available in the TwinsUK cohort. The † symbol denotes phenotypes measured in all participants. WCC — white blood cell count; DXA — Dual-energy X-ray absorptiometry; CV — cardiovascular; ECG — electrocardiogram; IBS — inflammatory bowel disease; OA – osteoarthritis; OB & GYN — obstetrics and genealogy; SF36

* adopted from http://www.twinsuk.ac.uk/data-access/cohortdata-description/

—The Short Form (36) Health Survey; GHQ — General Health Questionnaire; SES — socioeconomic status.

studies in the TwinsUK cohort were performed with the approval of the Guy's and St Thomas' Ethics Committee, and participants signed a written informed consent in accordance with the Declaration of Helsinki.

3.2.2. Omics data

Apart of the detailed phenotyping, a strength of the TwinsUK cohort is the availability of data from several 'omics' technologies (**Table 3.1.**). These include genome-wide scans of single nucleotide variants, next-generation sequencing, exome sequencing, epigenetic markers (both chip and sequencing), gene expression (arrays and RNA sequencing), telomere length measures, metabolomics (Metabolon and nuclear magnetic resonance (NMR) platforms), and gut flora microbiome.

Genotyping and imputation of single nucleotide polymorphisms (SNPs)*: The genotyping QC, imputation and post-imputation QC for the TwinsUK cohort were performed by Dr. Massimo Mangino, Dr. Pirro Hysi, Dr. Nicole Soranzo and Dr. Abhishek Nag. TwinsUK participants were genotyped using a combination of four Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2M-Duo). The probe intensity data was normalised and pooled for each of the three arrays separately (with 1M-Duo and 1.2M-Duo 1M pooled together). For each dataset the Illluminus calling algorithm was used to assign genotypes[466]. No calls were assigned if an individual's most likely genotype was called with less than a posterior probability threshold of 0.95. Validation of pooling was achieved via a visual inspection of 100 random, shared SNPs for overt batch effects. Finally, intensity cluster plots of significant SNPs were visually inspected for dispersion biased no calling, and/or erroneous genotype assignment. SNPs exhibiting any of these characteristics were discarded. The following exclusion criteria were applied to each of the three datasets separately. Samples were excluded if sample call rate was lower than 98%, heterozygosity across all SNPs was equal to or higher than 2 standard deviations (SD) from the sample mean; evidence of non-European ancestry as assessed by principal component analysis (PCA) in comparison to HapMap3 populations; observed pairwise identity by descent (IBD) probabilities suggestive of sample identity errors. Furthermore, misclassified monozygotic and dizygotic twins were corrected based on IBD probabilities. Variants were excluded if:

* includes small indels

the p-value* for deviation from Hardy-Weinberg Equilibrium (HWE)

Table 3.1. Summary of available Omics data in the TwinsUK cohort

Omics type	Platform	N$	Study
Genome-wide array genotyping	Illumina 317/610	5,654/6,300*	TwinsUK
	Illumina 1/1.2M-Duo	153	
Exome chip	Illumina HumanExome	4,080	TwinsUK
Whole-genome sequencing	Illumina 7x	1,800	UK10K
	Illumina 30x	2,000	TwinsUK/HLI
	Illumina 4x	325	GOT2D
Exome sequencing	Illumina 4x	925	GOT2D
Transcriptomics	Illumina HT12 array	850 (Blood/Adipose/Skin)	MuTHER
	RNA-seq	800 (4 tissues)/ 130 (3 time-points)	MuTHER/SystemoTwin
	Micro RNAseq	180	
Epigenetics	Illumina 450K array	900 xBlood/600 xAdipose/400 xSkin	Epitwin/MuTHER
	Illumina 27K array	170	Epitwin
	MeDIp Seq	4,000	
	Illumina EPIC	500	ESRC
Metabolomics	Biocrates	1,200	TwinsUK/HLI
	Metabolome	6,400/ 2,050 (3 time points)	
	NMR-based	2,050 (3 time points)	
	Faecal metabolomics	1,037	
	Salivary	2,723	
Microbiome	16S (x10K) from Gut	5,494 (2,731)	Flora
	16S salivary/urinary		
	Metagenomes	1,250	BGI/HLI
Proteomics	SOMAscan v3	300	
Glycomics	IgG/plasma	4,900/1,800	

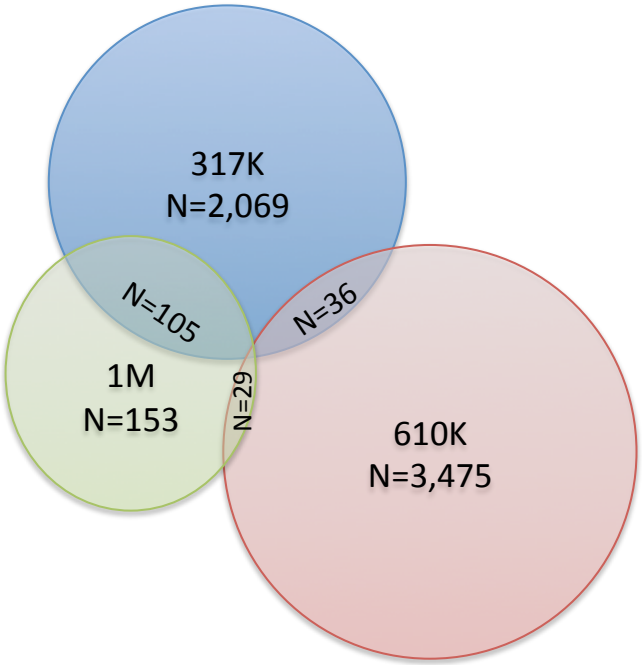
The table summarises the TwinsUK omics data gathered through various studies. The number (N) of samples measured and the platforms used for each omics type are listed. UK10K — UK 10,000 genomes project; GOT2D — Genetics of T2D consortium; HLI — Human Longevity Inc.; MuTHER — Multiple Tissue Human Expression Resource; ESRC — Economics and Social

* unrelated samples

Research Council; NMR — nuclear magnetic resonance; $ datasets with N>100 only; * untyped twins from MZ-twin pairs were added as copies of the typed co-twin's genotypes.

was less than $1x10^{-6}$; the minor allele frequency (MAF)* was less than 0.01; SNP call rate was less than 97% for SNPs with MAF≥5% or less than 99% for SNP with MAF between 1% and 5%. Alleles of all three datasets were aligned to HapMap3 forwards strand.

Figure 3.2. Summary of genotyping platforms used in the TwinsUK cohort

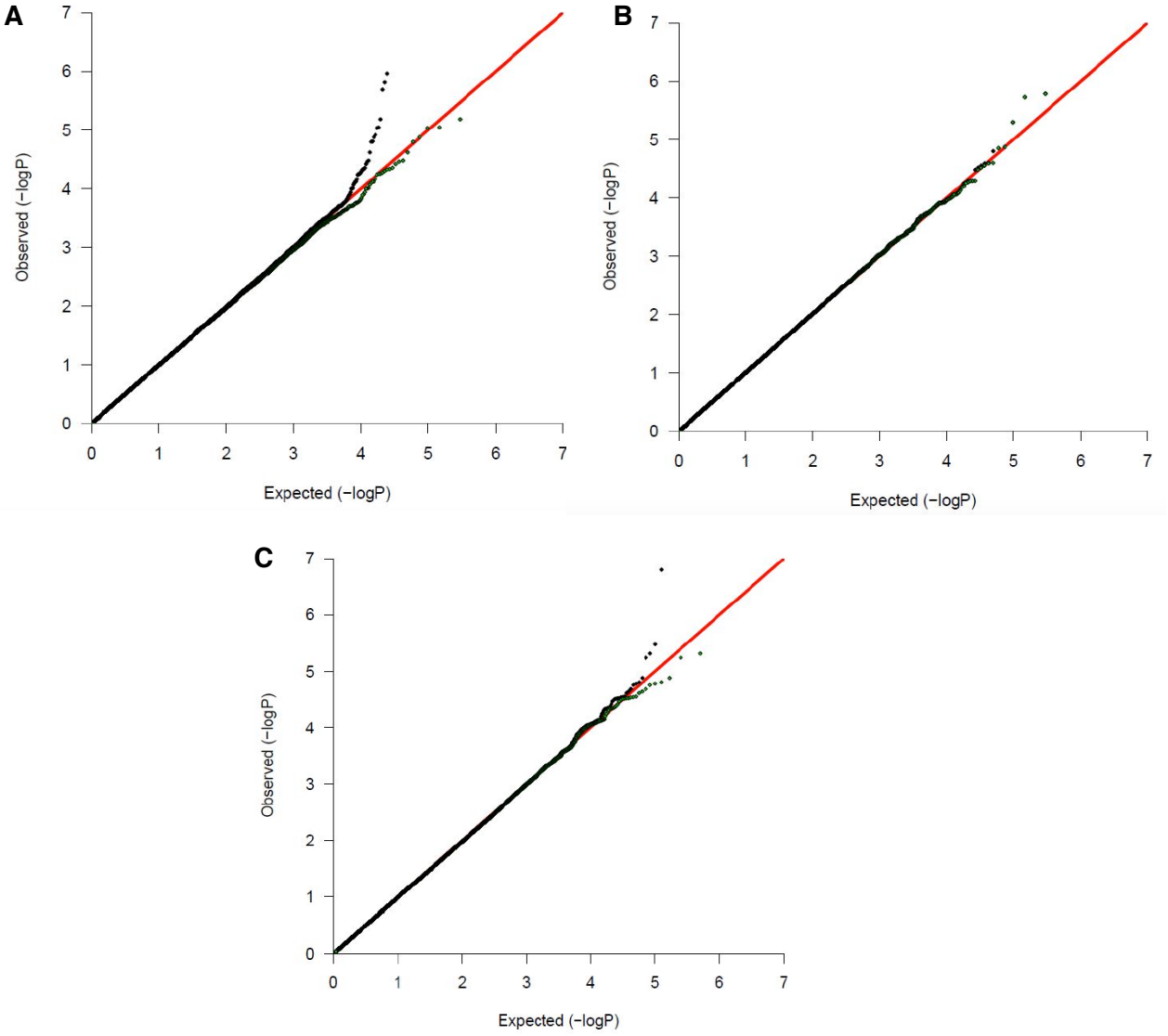


This figure shows the breakdown of people genotyped on each array in the final merged TwinsUK data-set of genotypes, as well as the overlap in number of samples between arrays. None of the samples were genotyped on all three arrays. Light blue circle — Illumina 317,000 SNPs array; Light red circle — Illumina 610,000 SNPs array; Light green circle — Illumina 1,000,000 Duo and 1,200,000 Duo arrays.

The data from the three platforms was merged keeping individuals typed at the largest number of SNPs when an individual was typed with two different arrays (**Figure 3.2.**). Prior to merging, the following pairwise comparisons among the three datasets were performed and, further SNPs and samples were excluded to avoid spurious genotyping effects. Samples and SNPs were excluded if: the concordance at duplicate samples was lower than 99%; the concordance at duplicate SNPs was less than 99%; SNP deviated from the null model (**Figure 3.3.**) in a logistic regression applied to all

pairwise dataset comparisons with array as the phenotype; the SNP HWE p-value* was $1x10^{-6}$; observed pairwise IBD probabilities were suggestive of sample identity errors. The final dataset contained 5697 individuals and 918,282 SNPs.

Figure 3.3. Quantile-quantile plots for pairwise array comparisons



This figure, courtesy of Dr. Manigino, shows the quantile-quantile (QQ) plots for models where genotyping array was used as binary phenotype and genetic variants were used as predictors. These analyses were done in unrelated individuals only. Black and green dots denote p-values prior and after the exclusion of discordant SNPs and individuals respectively. A) 317K versus 1M arrays; B) 317K versus 610K; C) 610K versus 1M arrays.

* unrelated samples

The TwinsUK cohort has been imputed against three different reference panels: HapMap3, 1000 Genomes (1000G) project, and the Haplotype Reference Consortium (HRC) panel. **In the present thesis only data imputed against the 1000 Genomes reference panel was used.** Therefore, the imputation and QC procedures only for this particular dataset will be presented here. The 918,282 SNPs that passed genotyping QC were imputed against the full (world) panel of individuals (Phase 3, version 2014) of the 1000G project. The "HRC/1KG Imputation Preparation and Checking Tool" (version 4.2.5, http://www.well.ox.ac.uk/~wrayner/tools/) was used to check input data for accuracy relative to 1000G inputs prior to imputation. This process identified errors in the original data, including incorrect REF/ALT designations, incorrect strand designations, extreme deviations from expected allele frequencies, and palindromic (A/T and G/C) SNPs with allele frequency near 0.5 that are often the source of imputation errors. The problematic variants identified were fixed or removed. Imputation was done on the Michigan Imputation Server* using Eagle version 2.3 for haplotype phasing† and IMPUTE2 for imputation[467,468]. No central post-imputation QC was performed. The steps I took to assure the SNPs I have used are well imputed are described in the next chapter (**Chapter 4**, Section: 4.3.3. Statistical analysis, Subsection: GWAS results QC).

Whole-genome sequencing data (provided by UK10K project): To determine the role of rare genetic variants in cataract formation, I have made use of the whole-genome sequencing data, which is also available for a subset of the TwinsUK cohort (N=1800‡). DNA from whole blood was whole-genome sequenced at low coverage (median coverage: 6x) as part of the UK10K project§. The 1800 individuals (one twin per pair) were selected by the TwinsUK data management team to maximise overlap with a number of existing genomic resources in addition to availability of a wide range of phenotypes and environmental exposure data. The sequencing was performed using Illumina Genome Analyzer and Illumina HiSeq 2000 platforms and all variants passed strict QC as discussed in detail elsewhere[469] and are briefly summarised below.

The sequencing was performed at both the Wellcome Trust Sanger Institute and the Beijing Genomics Institute. DNA (1-3µg) from PBMCs was sheared to 100-1000 bp (Covaris E210 or LE220 (Covaris, Woburn, MA, USA) and subjected to Illumina paired-

* https://imputationserver.sph.umich.edu/

† https://data.broadinstitute.org/alkesgroup/Eagle/

‡ post QC

§ http://www.uk10k.org/

end DNA library preparation[469]. Following size selection (300-500 bp insert size), DNA libraries were sequenced using the Illumina HiSeq platform as paired-end 100 base reads according to manufacturer's protocol[469]. Sequencing reads that failed QC were removed using the Illumina GA Pipeline[469]. Data and data generated at each of the two sequencing centres were aligned to the GRCh37 human reference separately by the respective centres[469]. Single nucleotide and insertion deletion (indels) polymorphism calls over all samples and all-sample and all-site genotype likelihood probabilities were generated using samtools and bcftools respectively[469]. Sites were filtered out if: their variant quality score log odds ratio values corresponded to less than 99.5% and 97% sensitivity for SNPs and indels respectively; had p-value for deviation from HWE of $1x10^{-6}$; showed batch effects between sequencing centres or participating cohorts[469]. Samples were filtered out if: there was high overall discordance to the array genotypes (> 3%); heterozygosity rate was more than 3SD from population mean suggesting possible contamination; the read depth per sample was low (below 4x mean read depth)[469].

Metabolomics single-time-point (N=6,400) and longitudinal measurements (N=2,050)[470,471]: In **Chapter 6** I make use of the TwinsUK metabolomics data. This data, in its fully QCed form, was generously shared with me by Dr. Jonas Zierer and Dr. Cristina Menni. Briefly, blood samples were collected after a minimum of six hours of fasting. After coagulation (4°C for 40 min), samples were centrifuged at 3,000 rpm for 10 minutes and the top clear yellow layer (either plasma or serum*) was aliquoted as 0.5 to 1 ml aliquots into 1.5 ml micro-centrifuge tubes. The tubes were stored at -45°C until metabolomics profiling. Untargeted metabolite profiles from blood were generated by Metabolon Inc. as follows. After being isolated, metabolites were separated using ultra-performance LC and GC, spiked with injection standards of fixed concentrations and then helium-loaded into an 5% phenyldimethyl silicone† MS column. The following mobile phases were used: water, 95% methanol containing 0.1% formic acid, or 6.5 mM ammonium bicarbonate. Samples were then vaporised (60°C to 340°C over for 17 minutes) and underwent electron ionisation, followed by acceleration, deflection and detection. The MS values were then compared to a reference library that used molecular weight, retention time and in-source fragment values in order to identify the metabolites as known chemical compounds. Metabolon Inc. provided relative concentrations of metabolites per sample. These were then quantile-normalised and any data points that were greater than 4 standard deviations from the mean were

* plasma without fibrinogens

† stationary phase

removed.

For a subset of TwinsUK participants, metabolites (from serum) were available from multiple collections (over a maximal 18-year period). The metabolites were measured again using the Metabolon platform but this time a newer detection platform was employed, that did not include GC and that afforded a better detection of lipid subspecies. Detailed description of the methods used is available elsewhere[471]. The effect of all technical covariates was accounted for prior the data release to us by Metabolon Inc. Missing relative concentrations were imputed using the minimum value of the run-day. The data was scaled by run-day median and inverse-normalised. Metabolites with more than 20% missing values were excluded.

Microbiome: In **Chapter 6** I also make use of the TwinsUK Microbiome data. This data was in its fully QCed form, was generously shared with me by Dr. Matt Jackson and Dr. Jordana Bell. Briefly, participants collected delivered a home-collected faecal sample (15mL conical tube, at 4°C for 1-2 days) to their annual TwinsUK clinical phenotyping visit. Samples were then frozen and stored at -80 C before being shipped to Cornell University on dry ice for sequencing. DNA was extracted from the 100mg of faecal sample (MoBio Power Soil® DNA isolation kit, MoBio Laboratories Ltd, Carlsbad,CA). The V4 hypervariable region of the 16S rRNA gene was PCRed using 515F and 806R primers[472] in duplicates (2.5U Easy-A high-fidelity enzyme, 1 x buffer (Stratagene, La Jolla, CA), 10-100 ng DNA template, and 0.05 mM of each primer). The following PCR conditions were used: 3 minutes of denaturation at 94°C; 25 cycles — 45 seconds of denaturation at 94°C C; annealing — 60 seconds at 50°C; extension — 90 seconds at 72°C; final extension — 10 minutes at 72°C. The PCR duplicates were combined, purified using the Mag-Bind® magnetic bead purification system (EZPure, Omega Bio-Tek, Norcross, GA) and quantified using the QuantiT PicoGreen dsDNA assay kit (Invitrogen, Carlsbad, CA). Aliquots were combined to yield a final concentration of approximately 15 ng per µl.

Paired-end sequencing was performed on the Illumina MiSeq platform at Cornell Biotechnology Resource Center Genomics Facility, following the manufacturer's protocols. The matched pair reads were joined using fastq-join and reads exceeding 275bp in length were discarded. Sequences containing unreadable barcodes, ambiguous bases or low quality reads (Phred quality score ≤25) were discarded, resulting in 81,634,331 quality-filtered sequences. Quantitative Insights Into Microbial Ecology (QIIME 1.7.0) was used to analyse the remaining sequences. Closed reference clustering was used to determine operational taxonomy units (OTUs) at 97% sequence identity against the Greengenes v13_5 taxonomy database. OTUs that were

observed in fewer than 25% of individuals were not considered for further study. This threshold was applied to focus on associations within abundant OTUs where data sparsity would be less influential on analyses. Of the total 9,840 OTUs, 16% passed this threshold (data sparsity) resulting in a final set of 2118 OTUs. These OTUs were further collapsed into taxonomies (N=560) at the family (N=124), genus (N=283 variables) and species (N=153) levels where only fully classified taxa were considered within each level.

OTU counts were converted to relative abundances and log10 transformed. Even though standard collection and sample processing were used PCA analyses revealed the following confounders: sequencing run, sequencing depth, who extracted the DNA, who loaded the DNA, and sample collection method. The transformed abundances were then used as the response in a model with those confounders as covariates. The resulting residuals from this model were used for any subsequent analysis.

3.2.3. Data collection in the TwinsUK cohort

As part of their PhD work, PhD students at the Department of Twin research and Genetic epidemiology are required to participate in clinical examination data collection. This is usually done for a day each week for 2 years. In addition to that, due to personnel shortage, I performed all data collection for the ophthalmologic component of the twin visit for a period of six months on full time basis. During this period, two new machines for eye measurements were introduced: Visionix VX120 — a combined all-in-one device for performing a whole variety of eye tests; and the Optovue iFusion system that performs optical coherence tomography scans and fundus retinal imaging. This required the development of new data collection protocols. The protocols I developed can be found in Appendix I. I was also responsible for the training of research nurses and students who later joined the department on how to use the machines and implement the protocols. The data collected during the 6 months period and later in my PhD contributed to both scientific articles and conference abstracts (see publications arising from this thesis: page 7 and Appendix II).

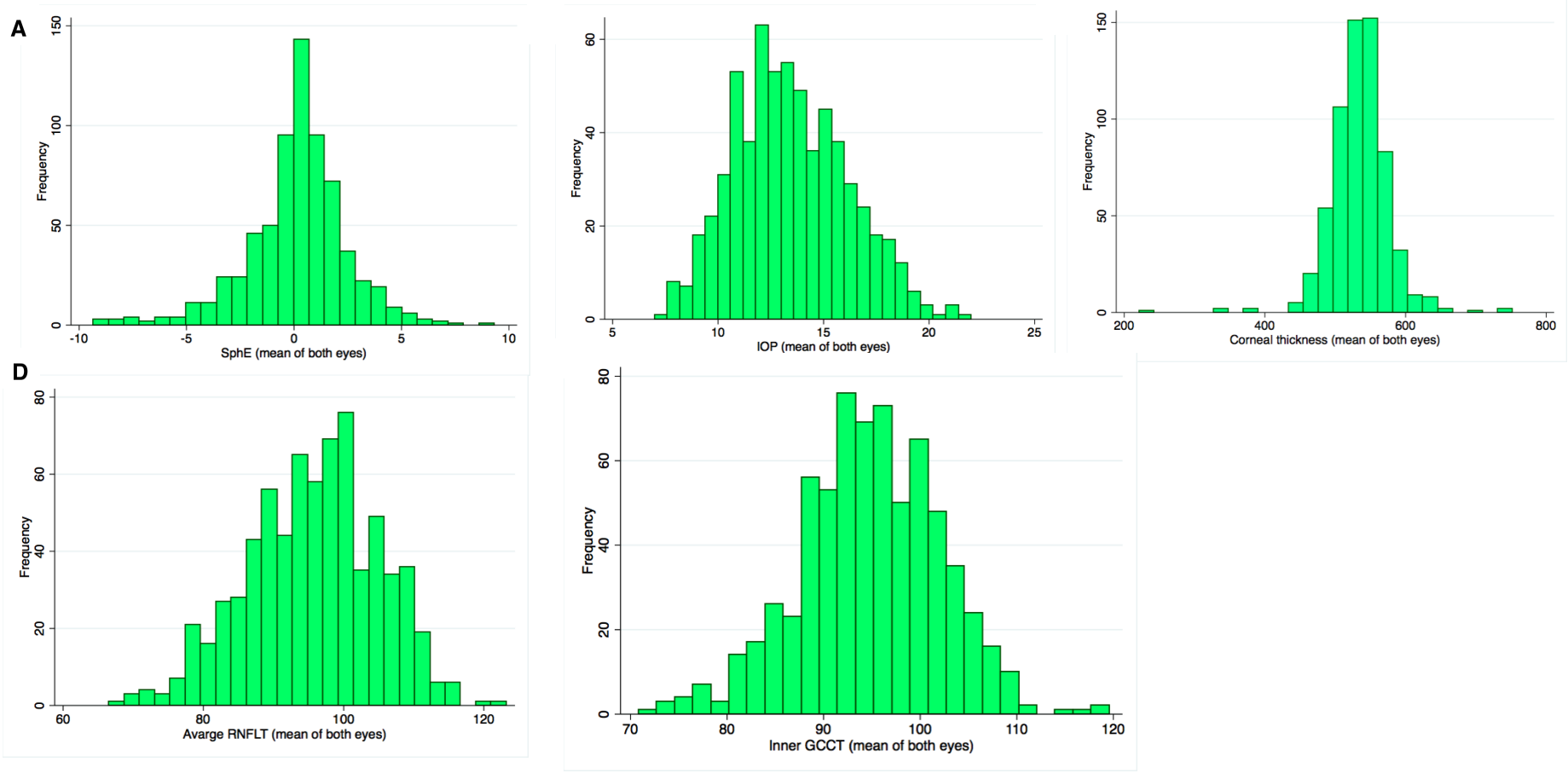
In the period from 01/09/2014 to 28/02/2015 I have measured 734 individuals (98.3% white British, 77.8% women, mean age (±SD) of 62.3±11.7). The key ophthalmic parameters measured in the white British individuals (N=708) are summarised in **Table 3.2.** and distribution plots are presented in **Figure 3.4**. As there was a high correlation between left and right eyes for all measurements (Pearson correlation coefficient >0.7), the mean of both eyes is presented in the table.

Table 3.2. Summary of ophthalmic parameters in TwinsUK participants measured between September 2014 and February 2015

Parameter	Men	Women
Number of individuals	52	571
SphE (mean±SD)	-0.24±0.35	0.17±0.93
IOP (mean±SD)	12.95±0.50	13.48±0.10
Corneal thickness (mean±SD)	520.33±9.14	536.88±1.63
Average RNFLT (mean±SD)	93.88±1.27	95.88±0.36
Superior RNFLT (mean±SD)	114.89±1.70	117.34±0.52
Nasal RNFLT (mean±SD)	73.89±1.56	74.15±0.42
Inferior RNFLT (mean±SD)	116.22±2.14	120.44±0.14
Temporal RNFLT (mean±SD)	70.52±1.14	71.61±0.33
inner GCCT (mean±SD)	94.2±1.06	94.90±0.29
superior GCCT (mean±SD)	93.40±0.97	94.30±0.30
inferior GCCT (mean±SD)	95.00±1.07	95.48±0.31

This table summaries key ophthalmic parameters in the 708 white British individuals measured by me between September 2014 and February 2015. The mean of both eyes is presented. Values for men and women are given separately. SD — standard deviation; SphE — spherical equivalent; IOP — intraocular pressure; RNFLT — retinal nerve fibre layer thickness; GCCT — ganglion cell complex thickness

Figure 3.4. Distribution plots of eye parameters



This figure shows the distribution of the mean of both eyes for some of the key eye parameters in the subsets of TwinsUK individuals measured by me in the period 1st September 2014 to 28th of February 2015. A) Spherical equivalent (SphE); B) Intraocular pressure (IOP); C) Corneal thickness; D) Average retinal nerve fibre layer thickness (RNFLT); E) Inner ganglion cell complex thickness (GCCT)

3.3. The Rotterdam Study

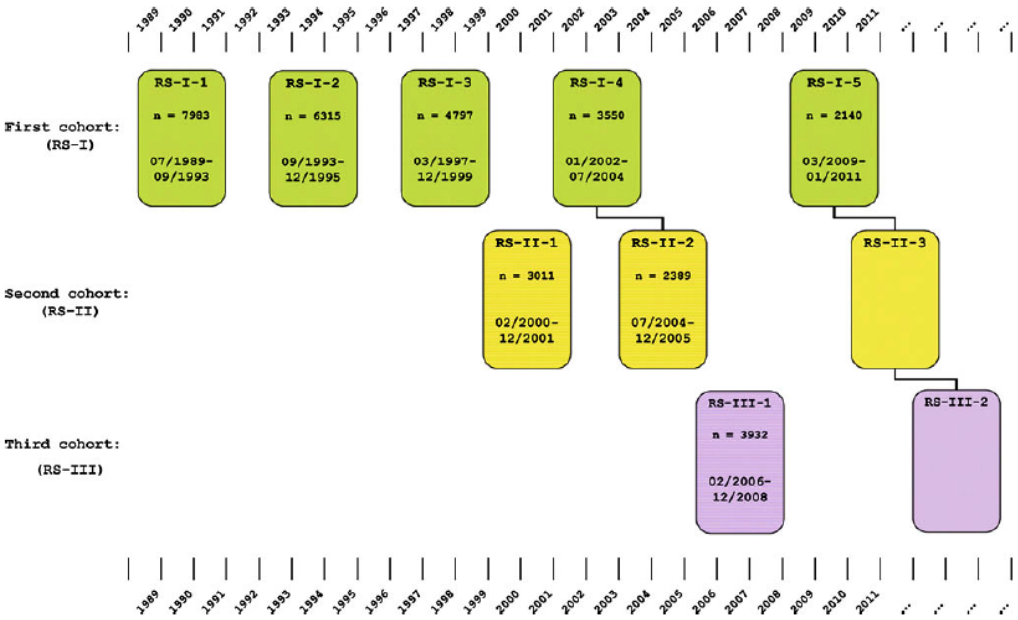
3.3.1. Cohort description

The Rotterdam Study (RS) is a Dutch prospective population based study (N= 14,926) focusing on cardiovascular, neurological, ophthalmological, endocrine and psychiatric diseases in the elderly[473-475]. The study sampled white Dutch individuals (100%) from both sexes (60% women) living in Ommoord (a suburb of the city of Rotterdam) during three independent waves of enrolment (**Figure 3.5.**), resulting in three cohorts (RSI, RSII, and RSIII)[473]. The participants in each cohort were examined at multiple time points at regular intervals from 1991 to present (eg. baseline: RSI-1, RS2-1, RS3-1; first follow-up: RSI-2; RSII-2; RSIII-2; etc.). All individuals were first interviewed at home (2 h) and then had an extensive set of examinations (a total of 5 h) including imaging (of heart, blood vessels, eyes, skeleton and later brain) and on collecting biospecimens that enabled further in-depth molecular and genetic analyses. All of the participants in the Rotterdam study have signed written informed consent and the Medical Ethics Committee of Erasmus Medical Centre has approved the study.

At baseline (RDSI; 1989-1993), 7,983 individuals (78% response rate) were enrolled (age range: 55-106). The Rotterdam study II started enrolling in 2000 (N=3,011; 67% response rate; age 55 years or older), while the Rotterdam study III, a further extension of the cohort from 2006 onwards, enrolled slightly younger individuals (N=3,932, aged 45–54 years, 65% response rate)[473-475]. Since summer of 2016, another extension RSIV has started that includes all participants aged 40 years and over[475]. The recruitment of this extension is expected to be completed in 2019 and yield around 4000 new participants[475].

As part of the ophthalmic examination, assessment of cataract was carried out only in a subset of the Rotterdam Study cohorts (Rotterdam study I: visits 3 and 4, Rotterdam study II visit 2, and Rotterdam study III visit 1). **In the current thesis only data from the Rotterdam study I visit 3 was used. From now on throughout this thesis the RSI-3 is referred as RSI-III.**

Figure 3.5.Diagram of examination cycles of the Rotterdam Study



This figure is adapted from Hofman *et al.*, Eur J Epidemiol 2011[476] and show the examination cycles of the Rotterdam Study.

Similarly to the TwinsUK cohort, the Rotterdam study has omics data available (genome-wide (array and sequencing): genotyping, transcriptomics, and methylation; as well as metabolomics, proteomics and microbiome data[475]. In this thesis, I have made use of only the genome-wide array, that was subsequently imputed to the 1000G reference panel. The genotyping and imputation procedures used in the Rotterdam study are shown below. This data has been provided to me by Prof. Dr. Caroline Klaver and her team who are responsible for the ophthalmologic studies within the Rotterdam Study.

Genotyping and imputation of single nucleotide polymorphisms (SNPs)*:
DNA was extracted from blood leucocytes according to standard procedures. Genotyping of SNPs was performed using the Human 610 Quad Arrays Illumina (RSI-III). Samples with low call rate (< 97.5%), with excess autosomal heterozygosity (> 0.336), or with sex-mismatch were excluded, as were outliers identified by the identity-by-state (IBS) clustering analysis (outliers were defined as being > 3SD from population mean or having IBS probabilities > 97%). A set of genotyped input SNPs (N = 530,683) with call rate > 98%, with MAF > 0.01, and with HWE p-value >$1x10^{-6}$ was used for imputation. Imputation was conducted against the in European Ancestry panel

* includes small indels

of individuals (Phase 1, version 2012) of the 1000G project using an algorithm implemented minimac, after phasing using MACH[468,477]. As with TwinsUK, no central post-imputation QC was performed. The steps I took to assure the SNPs I have used are well imputed are described in the next chapter (**Chapter 4**, Section: 4.3.3. Statistical analysis, Subsection: GWAS results QC).

3.3.2. Data collection in the Rotterdam Study

As part of the third follow-up of the Rotterdam Study (January 1998 to December 1999), 3,600 participant attended an eye test. During the eye test lens images were taken (N=15,599) for both eyes using an analogue camera. Nuclear cataracts were assessed using Slit Lamp/Scheimpflug (1 image per eye), while cortical and PSC cataracts (2 images per eye) were assessed using Neitz CT-R camera (Kowa Optical, Torrance, CA). In the latter case, one of the images was taken at higher depth to permit better focus on the posterior lens, adjacent to the posterior lens capsule. After being taken the images were developed, mounted on slides and stored. At the time the quality of the images was not evaluated and cataracts were not graded. As part of my PhD I had the task to scan/digitise all those images, evaluate their quality and perform cataract grading. The grading procedures I used will be detailed later in this thesis (**Chapter 4**; Section 4.3.2.Phenotyping). Here I will focus on the procedures followed when scanning the images and on the quality score I developed in order to determine whether the images were of high enough quality to be reliably graded.

Image scanning: On each slide the participant's study number, the date on which the photo was taken and the participants sex were noted. This information was used to find the participants in the central database and to create a barcode that the scanner could read. The barcode included information of the participant's study number and sex, the date of visit, and the type of photo (Scheimpflug/retroillumination) taken. The barcodes were read in a particular order that followed the order in which the slides were mounted on the scanner feeder trays. The slide with lens images (3 x eye when possible*) were then scanned on a Hasselblad Flextight X5 Scanner using the following standard settings: 16MB RGB standard; 36x24 full size frame; auto levels; 26% scanning resolution; output PPI=3000, output file format = .tiff. The images were automatically named using the information from the barcode. Finally, images were converted to grey scale using Apple Automator.

* 1) some images were extremely dark, thus the scanner could not detected that a slide was present; 2) a few individuals had one or more of their lens photos missing

Table 3.3. Quality scores for RSI-III cataract images

Parameter	Points	Meaning	N
Images of the lens nucleus			
Brightness: amount of light on the lens nucleus	0	normal	3,169
	1	too bright	1,745
	2	too dark	569
Position: lens nucleus is visible	0	right	3,806
	1	wrong	1,683
Focus: is the lens in focus	0	yes	4,193
	1	no	1,450
Reflex: visible and in the middle of anterior chamber	0	yes	3,703
	1	no	1,240
Dilation of the eye	0	full	4,888
	1	partial or minimal	655
Lid: obscures partially or fully the nucleus	0	no	5,184
	1	yes	639
Other problems leading obscured lens nucleus: reagents on slide; artefact on slide; lens is not visible; double exposure	0	no	5,540
	1	yes	3
Overall quality (total score):	**0**	**good**	**5,216**
	>0	**issues present**	**327**
Retroillumination images			
Dilation of the eye	0	full	3,422
	1	partial or minimal	2,122
Lid: obscures partially or fully the lens	0	no	4,936
	1	yes	590
Brightness: amount of light on the lens	0	normal	5,236
	1	too bright	308
Position: the total lens surface is visible	0	right	5,540
	1	wrong	4
Focus: is the lens in focus	0	yes	5,521
	1	no	23
Overall quality (total score):	**0**	**good**	**2264**
	>0	**issues present**	**3280**

This table summarises the parameters used to assign a quality score to an RSI-III lens image. The number of eyes (N) that fall in each category are also listed. Eye where the lens had been surgically removed are included in the N presented here.

Image quality assessment: In all cases, both eyes were dilated prior to taking the photograph to permit the whole lens to be visible. When assessing the images, however, I discovered that no standard conditions were used. For example sometimes

images were taken in full-light conditions or before the pupil was fully dilated. In addition photos of aphakic and pseudophakic eyes were also taken. These issues did not substantially affect the grading of cortical and PSC cataracts but they were problematic when assigning nuclear scores because the use of objective grading systems, such as the CNDS grading system, requires standard lighting and positioning of the lens. Therefore, I crated a scoring system (**Table 3.3.**) and scored each photo using this system. Images with a total score higher than 0 were deemed unacceptable and were not used in any subsequent analysis. In addition, prior to the analysis, I excluded individuals with bilateral cataract surgery or with mislabeled photos*.

3.4. International Cataract Genetics Consortium

3.4.1. The Age-Related Eye Disease Study (AREDS)

Cohort description: AREDS is a USA-based long-term prospective multi-centre study of the clinical course of age-related macular degeneration (AMD) and age-related cataract, that included a clinical trial of high-dose vitamin supplements for AMD and cataract[382,420]. After signing informed consent according to the Declaration of Helsinki, the AREDS participants (N=4,757, age at enrolment (1992-1998): 55 to 80 years; 57% women; 96.7% white) were enrolled on the basis of fundus photographs graded by a central reading centre, best-corrected visual acuity and ophthalmologic evaluations and had to be free of any illness or condition that would make long-term follow-up or compliance with study medications unlikely or difficult. Lens images were taken using slit lamp and cataract were graded using the AREDS system[478].

Genotyping and Imputation: All AREDS participants were genotyped at the Centre for Inherited Disease Research. For AREDS1a-1b three chips were used for this genotyping: Affymetrix 100K, Illumina 100K and Illumina 300K. SNPs were abstracted from each of the chips and genotypes on more than one chip were checked to ensure that the calls were the same. Individuals not of European descent were removed. For AREDS 1c, genotyping of SNPs was performed using the Illumina HumanOmni2.5-4v1_B chip array. **For the analysis presented in Chapter 5 only individuals genotyped on the HumanOmni2.5-4v1_B chip array were used.** For all studies, samples with low call rate (<98%), with low mean confidence scores over all non-missing genotypes, with chromosome anomalies, or with sex-mismatch were

* the information noted on the slide was different than that in the central data base

excluded. No samples exhibited excess heterozygosity rates (1.5 interquartile ranges above or below the upper/lower quartile ranges). Cryptic relatedness was detected by estimating IBD sharing and kinship coefficients among all possible pairs and one member of each pair exhibiting a sibling or closer relationship was dropped from the analysis. SNPs were dropped from the analysis if they exhibited more than 1 blind duplicate error, more than 1 HapMap control error or more than 1 error in HapMap control trios, a genotype call rate < 99%, minor allele frequency < 0.01, or HWE p-value < 10^{-4}. No batch effect, sex-specific differences in allelic frequency (> 0.2) or heterozygosity (> 0.3) were detected. Imputation was performed with IMPUTE2 against 1000G world samples from phase I integrated variant set (v3, release March 2012). Imputed variants were aligned the to plus strand of NCBI build 37, for each imputed SNP a measure of imputation quality were calculated (info score) was calculated. No central post-imputation QC was performed. The steps I took to assure the SNPs I have used are well imputed are described in the next chapter (**Chapter 4**, Section: 4.3.3. Statistical analysis, Subsection: GWAS results QC).

3.4.2.The Beaver Dam Eye Study (BDES)

Cohort description: BDES is a population based study of age related ocular disorders. All individuals (N = 5925) who were between 43 and 84 years of age as determined by a private census of the township, and resided in the city or township of Beaver Dam, Wisconsin, USA, in 1987-8, were eligible to take part in the study[479]. Of the 5925 eligible people identified, 4926 (83.1% response rate) were examined at baseline (1988-90). These people were white (99%) and 56% were female. The cohort was re-examined at 5- (N = 3722), 10- (N = 2962), 15- (N = 2375), and 20-year (N = 1913) follow-ups. Physical examinations and lens photography were conducted both at baseline and at follow-up. Medical histories and information on demographic and behavioral characteristics were ascertained at baseline, using a standardized questionnaire administered in conjunction with the physical examinations. Lens photographs were taken using a Topcon SL5 Slit Lamp camera (Paramus, New Jersey) with specially designed fixation targets for photographing the nuclear region of the lens. In the BDES cataracts were graded using the original Wisconsin grading system[39].

Genotyping and Imputation: BDES samples were genotyped using Illumina HumanExome BeadChip (San Diego, CA, USA) at the Genetic Resources Core Facility at Johns Hopkins Institute of Genetic Medicine[480]. Genotypes were called using the Illumina GenTrain clustering algorithm in GenomeStudio. All samples had a call rate >

98% but individuals with sex inconsistencies (N = 15) and Mendelian errors (N = 2) were excluded. Cryptic relatedness and unexpected duplicates were determined by identifying pairs of individuals with IBD sharing > 20%, representing first and second-degree relatives. The individual from each relative pair with the lower quality score (n = 85) was excluded. Six individuals did not cluster with HapMap European ancestry controls (CEU), and were excluded. Of the 242,901 genotyped variants, variants that fell in the following categories were excluded: non-autosomal variants (N = 5,465), SNPs with a call rate<98% (N = 6,121), and monomorphic variants (N = 138,061)[480]. Variants were aligned the to plus strand of NCBI build 37 and imputed against the 1000G world samples from phase I integrated variant set (v3, release March 2012); using an algorithm implemented minimac, after phasing using MACH[468,477]. For the analysis presented in this thesis I have used this cohort as a replication cohort, focusing only on the subsets of variant that were taken forwards for replication as part of the meta-analysis of GWAS of nuclear cataract project (see **Chapter 5**).

3.4.3. The Beijing Eye Study (BES)

Cohort description: BES is a population-based cross-sectional study of Chinese individuals residing in rural areas (Yufa, Daxing District) south of and in urban areas north of Beijing (Haidian District)[239]. The Medical Ethics Committee of the Beijing Tongren Hospital approved the study protocol and all participants gave informed consent. At the baseline survey in 2001, a total of 5,324 individuals (40 years of age or older) were eligible to participate in the study, of which 4,439 individuals (83.4% response rate) were recruited. In 2006, all survey participants were re-invited and 3,251 participants (73.2% response rate) were recruited[239]. Participants underwent a series of examinations including: refractometry; pneumotonometry; biomicroscopy assisted by slit-lamp; optical coherence tomography of the anterior segment; photography of the cornea, lens, optic disc, macula and fundus; blood sampling for laboratory tests; blood pressure measurements, as well as determinations of anthropomorphic parameters[239]. They were also asked to complete a questionnaire which included questions on socioeconomic parameters, and awareness and treatment of ocular and general diseases.

In BES, the degree of nuclear cataract was assessed in 6 grades using the AREDS grading system, which is an extension of the Wisconsin System[39,478]. Digital photographs of the lens were obtained using the slit lamp. A single grader under the

supervision of senior ophthalmologists and researchers performed the lens grading. The nuclear cataract grading from the 2006 follow-up study was used[243].

Genotyping and Imputation: DNA was extracted from blood leucocytes according to standard procedures. Genotyping was performed using Illumina Human610-Quad BeadChip and Illumina OmniExpress arrays. Samples were excluded based on: call-rates of < 95% and excessive heterozygosity; cryptic relatedness; gender discrepancies; and discordant ethnic memberships as defined using PCA analysis .SNPs were excluded based on high rates of missingness (> 5%); monomorphism; or gross departure from HWE p-value < $1x10^{-6}$.

3.4.4. The Blue Mountains Eye Study (BMES)

Cohort description: BMES is a population-based survey of vision and common eye diseases in the Blue Mountains, west of Sydney, Australia[481]. Following a door-to-door census of the region, all permanent residents aged 49 years or older were invited to take a detailed eye examination[481]. From a total of 4,433 eligible residents, 3,654 (82.4%) gave informed consent, attended the baseline eye examination and were interviewed between 1992 and 1994. During 1997-99 (BMES II A); 2,335 participants (75.1% of survivors) returned for examinations after 5 years. During 1999-2000, 1,174 (85.2%) new participants took part in an Extension Study of the BMES (BMES IIB). At each examination, a detailed assessment of eye disease and other general health measures was conducted. Participants were also asked to attend fasting blood tests after each examination (except 15-year) and complete a detailed questionnaire about the types of food they consumed. At the follow-up exams, a test of memory and cognition (Mini Mental State Examination), questions about quality of life (Short Form 36) and visual functioning were also conducted.

Eye conditions were assessed at each examination by taking a series of photographs of the eye. Photographs of the retina were used to assess the presence of diseases like macular degeneration or diabetic eye damage. Stereo photographs of the optic nerve were taken to assess changes that, together with an automated test of the field of vision, indicate the presence of glaucoma. Two types of photographs were taken of the lens inside the eye to grade the presence of the different types of cataract. All of these photographs were graded using the Wisconsin grading system[39].

Genotyping and Imputation: Participants of the BMES were genotyped using the Illumina Human 670-Quadv1 custom genotyping array at the Wellcome Trust Sanger Institute, Cambridge as part of WTCCC2, and 2,761 had genotyping data

available. QC procedures in the WTCCC2 excluded SNPs with Fisher information measure < 0.98, genotype call rate < 0.95, MAF < 0.01 or HWE p-value < $1x10^{-6}$. Samples were excluded if they were identified as outliers on call rate, heterozygosity, ancestry and average probe intensity based on a Bayesian clustering algorithm. Samples were also removed if they exhibited discrepancies between inferred and recorded sex or cryptic relatedness with other WTCCC2 samples (pairwise IBD > 0.05). Imputation was than carried out against the European Ancestry panel of individuals (Phase 1, version 2012) of the 1000G project using IMPUTE. No central post-imputation QC was performed. The steps I took to assure the SNPs I have used are well imputed are described in the next chapter (**Chapter 4**, Section: 4.3.3. Statistical analysis, Subsection: GWAS results QC).

3.4.5.The India Eye Study (INDEYE)

Cohort description: INDEYE is a population-based study of people aged 60 years and older from a mix of rural and urban areas in North India (Gurgaon district, Haryana state) and South India (Pondicherry union territory and Cuddalore district in Tamil Nadu). The INDEYE study is aimed at estimating the age- and sex-specific prevalence of early and late AMD and of lens opacities, and at investigating associations of these conditions with tobacco use, exposure to biomass cooking fuels, outdoor work, and dietary factors[227,482]. Based on a randomly selected clusters where 8% of the total population per cluster would be aged 60 years and older, a total of 7,518 people (3586 from 29 cluster in North India and 3932 from 30 cluster in South India) were invited to take part in the study (2005-2007). In both locations the mean age of participants was 62 years and 52% were women. Participants were examined by an ophthalmologist either at their homes or during a hospital or ophthalmic centre visit. Informed written consent was obtained from all participants before enrolment. Information was read to people who were illiterate in the presence of a local witness, and a thumb impression of the participant signified assent. The study complied with the guidelines in the Declaration of Helsinki, and ethics approval was received from the Research Ethics Committees of the All India Institute of Medical Sciences, Aravind Eye Hospital, London School of Hygiene and Tropical Medicine, Queens University Belfast, and the Indian Council for Medical Research.

Genotyping and Imputation: In the current thesis only samples from the South India were used. The INDEYE has no genome-wide genotype data and was used only for replication purposes. Variants were genotyped using TaqMan.

3.4.6. Singapore Cohorts (SCES, SiMES, SINDI)

Detailed information for the Singapore cohorts is provided below. For the 3 cohorts in Singapore, ethics approval was obtained from the SingHealth Centralised Institutional Review Board. All study participants were provided with written informed consent in adherence to the Declaration of Helsinki.

Cohort description: The Singapore Chinese Eye Study (SCES-610 and SCES-OmniExpress) is a population-based cross-sectional epidemiological study on eye diseases in Chinese (age range: 40-80+; 50.6% were female) residing in Singapore. Using age-stratified random sampling strategy, 4605 ethnic Chinese residents in the southwestern part of Singapore were eligible from the sampling frame (N = 6,752), of which 3,353 (72.8% response rate) participants were recruited between February 2009 and December 2012[483].

The Singapore Malay Eye Study (SiMES) is a population-based cross-sectional epidemiological study on eye diseases for Malays aged between 40 and 80 years old residing in Singapore[484]. Details of the SiMES design, sampling plan and methodology have been reported elsewhere. In brief, between August 2004 and June 2006, a total of 4,168 Malay residents in the south-western part of Singapore were identified through age-stratified random sampling and were invited to participate in the study, of which 3,280 (78.7% response rate, 51.8% women) underwent a detailed ocular examination.

The Singapore Indian Eye Study (SINDI) is a population-based cross-sectional epidemiological study on eye diseases for ethnic Indians aged between 40 and 80+ years old residing in Singapore. The study was conducted between March 2007 and December 2009. Using age-stratified random sampling, 4,497 ethnic Indian residents in the south-western part of Singapore were eligible from the sampling frame (N = 6,350), of which 3,400 participants (75.6% response rate; age: 57.8 ± 10.1 years; 49.5% women) were recruited[483].

In the Singapore cohorts, the severity of nuclear cataract was assessed using the Wisconsin Cataract Grading System, based on lens photographs and followed a decimalized system (decimal scores from 0.1 to 5.0)[39]. In brief, lens photographs were taken using a digital slit-lamp camera (model DC-1 with DF-21 flash attachment; Topcon, Tokyo, Japan) and grading was performed through comparison with four standard photographs at the University of Sydney[42].

Genotyping and Imputation: For all studies, the Illumina Human610-Quad Beadchips (Illumina Inc.) were used for genotyping[27]. A first round of SNP QC was performed to obtain a cleaned set of genotypes for sample QC by excluding SNPs with:

high missingness (> 5%); gross departure from HWE p-value < 10^{-6}; and were monomorphic[27]. Samples were excluded based on: call-rates of less than 95% and excessive heterozygosity; cryptic relatedness; gender discrepancies; and discordant ethnic memberships as defined using PCA analysis[27]. A final round of SNP QC was applied to the remaining samples on autosomal SNPs using the same three criteria earlier, and an additional restriction of minor allele frequency > 5%. Genome-wide imputation of SNPs was performed on post-QC set of genotype data, using minimac with 1000G Phase 1 version 3 release (world panel). The same SNP QC procedure was performed on the imputed SNPs, except for the MAF threshold which was relaxed to 1%.

3.4.7. South London Case Control Study (SLCCS)

Cohort description: SLCCS is a case-control study, aimed at recruiting a replication sample for studies of the genetics of age-related eye diseases in South London. Study participants were recruited from the Princess Royal University Hospital, Orpington, and St Thomas' Hospital in London. Cataract cases were defined as having significant nuclear cataract (grade >= 3 on the LOCS III clinical grading system), controls were individuals without significant cataract (as assessed by treating physician, usually without LOCS III grades) and attending the eye department for other reasons (e.g. patients with glaucoma, age-related macular degeneration or other conditions). The SLCCS had no exclusion criteria on the grounds of age, gender, or ethnic origin, but for this study subjects identified as being of Northern European origin on principal component analyses were included.

Genotyping and Imputation: The SLCCS study was genotyped using Illumina OmniExpress Exome Array and imputed against the 1000G world samples from phase III integrated variant set (2014 release).The genotyping and imputation procedures used were exactly the same as the ones used for the TwinsUK cohort (this chapter; Section 3.2.2. Omics data). For the analysis presented in this thesis I have used this cohort as a replication cohort and I have not performed any genome-wide analysis; focusing only on the subsets of variant that were taken forwards for replication as part of the meta-analysis of GWAS of nuclear cataract project (see **Chapter 5**).

Chapter 4. Different genetic variants are associated with nuclear and cortical cataract

4.1. Overview

The aim of this chapter is two address the first three objectives of the current thesis:

1. Explore whether nuclear and cortical cataract share genetic risk factors
2. Identify common genetic variants associated with ARC by performing genome-wide association studies
3. Identify whether low frequency and rare genetic variants from whole-genome sequencing data are associated with ARC

In **Chapter 1** (section 1.7. Epidemiology and risk factors), the environmental risk factors associated with ARC have been discussed in detail. The different studies presented show that, apart from age, nuclear and cortical cataracts are unlikely to share much in terms of associated epidemiological risk factors. However, there is a general lack of studies focusing on answering questions related to causation, thus making it difficult to know whether nuclear and cortical cataract are different forms of the same disease or two totally unrelated diseases that happen to manifest in the same organ*. Answering this question is important not only from a mechanistic perspective but also because it can elucidate the need for different prevention or treatment strategies for the two different types of cataract. In terms of genetic studies, if the two types of cataract would share a substantial amount of genetic risk, then easily obtainable phenotypes such as “any cataract” or cataract surgery could be used in genetic analysis as opposed to a more detailed and time consuming grading of cataract by type. In **Chapter 4**, I explore whether nuclear and cortical cataract share genetic risk factors employing methods that use either phenotypic data only (classical heritability calculation) or genotypic data (amount of trait-shared phenotypic variance explained by the effect of a set of genetic variants).

Two of the analyses employed in answering the question of shared heritability (LD-score regression and polygenic risk scores) require the use of results from genome-wide association analysis (GWAS). GWAS are a standard tool for discovery of genetic loci associated with common diseases. However, as pointed out in **Chapter 1**

* PSC cataract has a low prevalence, and a population-based sample of 2,556 TwinsUK participants with cataract data (mean age of 62) lacks the number of affected individuals needed for meaningful analysis. Therefore, the thesis focuses on the two most common types of cataract — nuclear and cortical cataracts.

(section 1.6. Genetics of age-related cataract), apart of linkage analysis and candidate gene/variants studies, GWAS had not been employed in cataract genetics at the time the work on this thesis commenced (October 2013). My aim was, therefore, to explore the role common variants play in the pathogenesis of cataract by performing GWAS in two European cohorts — TwinsUK and the Rotterdam Study). Moreover, the first ever published GWAS of cataract studied Asian individuals only[27]. Therefore, in this chapter the findings from the first two GWAS in European ancestry cohorts will be discussed while in the following chapter (**Chapter 5**) I will focus on meta-analysis of GWAS.

In addition to genome-wide array data, the TwinsUK cohort also (**Table 3.3.**) has genome-wide sequencing data available. This data is used here to address the question from objective 3, namely whether low frequency and rare genetic variants, not surveyed by classical GWAS analysis, play a role in cataract formation.

4.2. Introduction

In this chapter three different analysis strategies have been employed to answer the above stated aims. These strategies are introduced in more detail below.

4.2.1. Shared heritability: definition and estimates

One way to answer the question of whether two traits share genetic risk factors is to calculate the amount of shared heritability between the two traits where heritability is defined as the proportion of phenotypic variance (V_P) that is explained by genetic factors. In this chapter, shared heritability is estimated both by using phenotypic data only and by using the estimates of the effect of SNPs on the traits. The first case is known as classic heritability modelling where the variance and covariance observed between related individuals is used to partition V_P of a trait as follows:

$$V_P = V_A + V_D + V_C + V_E,$$

where: V_A, also known as narrow sense heritability (h^2), is the variance explained by additive genetic factors (the combined direct effect of all genetic variants); V_D is the variance explained by the effect of alleles on the trait by allele interaction at the same (dominance (D)) or different loci (epistasis); the sum of V_A and V_D is known as broad sense heritability (H^2); V_C is the variance explained by common (C) environmental factors — all environmental exposures shared between individuals of the same population; V_E is the variance explained by unique (E) environmental factors that apply only to single individuals in the population (V_E includes measurement error)[485]. In case

of two- or multi-trait analysis V_P is the phenotypic variance shared between traits and V_A is the proportion of shared phenotypic variance that is explained by shared genetic factors between the traits.

The most powerful model for calculating heritability uses twins and assumes that monozygotic (MZ) twins share all of their genetic and common environmental factors and that any discordance is caused by unique environmental exposures. Dizygotic (DZ) twins, however, share on average half of their alleles and thus half of the genetic factors, but all common environmental factors. As a result, DZ discordance is attributed to the unshared half of their genetic components and the unique environmental exposures.

Heritability is a population-specific but not a constant measure (can change with time) that can predict the statistical power of gene mapping studies as high heritability reflects a high correlation between phenotype and genotype; and high shared heritability reflects high sharing of genetic factors[485]. However, heritability makes no claim about the architecture of a trait (number of associated loci, strength of effect the associated loci) and it makes certain assumptions (no or minimal gene-by-environment interactions, equal sharing of V_C for both MZ and DZ twin pairs, no trait-related selective assortment) that might not hold true in all cases[485,486].

Heritability, as well as shared heritability, can also be estimated using unrelated individuals but in this case both genetic and phenotypic data are required. One way to do that is to fit a penalised linear mixed effect model that treats the effect of SNPs as random and the variance explained by all the SNPs together is estimated:

$$Y_i = \mu + \sum_{i=1}^{m} z_{ij}\, u_i + \varepsilon_i$$

where: y is the phenotype, μ is the mean term, z is a allele frequency dependant penalty for m number of SNPs, u is the additive effect at each SNP and ε is the error term[487]. This method is implemented in the GCTA software[488]. Another method, also applied in this chapter, is the LD-score regression which regresses the GWAS summary statistics against the LD-score in order to calculate heritability:

$$E\,(X^2 \mid \Sigma_k r^2_{kj}) = N\, h^2\, \Sigma_k r^2_{kj} / M + Na + 1$$

where: X^2 is the summary statistics given the LD-score ($\Sigma_k r^2_{kj}$); N and M are the number of individuals SNPs respectively and h^2 is the heritability[489].

Yet another method that can be used to explore sharing of genetic factors is to

calculate a polygenic risk score for a trait and use it to predict another trait. This score is a sum of trait-associated alleles across many genetic loci, weighted by their effect size. The score is calculated using GWAS results or a meta-analysis of GWAS summary results as an initial training sample, and the markers are ranked by their evidence for association, usually their p-values[490]. An independent replication sample is then analysed by constructing, for each subject, a polygenic score consisting of the weighted sum of its trait-associated alleles, for some subset of top ranking markers[490].

4.2.2. Genome-wide association analysis

When studying the genetics of moderately heritable common multifactorial diseases in population-based cohorts the method of choice is GWAS. In GWAS a dense set of single nucleotide polymorphisms (SNPs) across the genome is genotyped to survey the common genetic variation for a role in disease or to identify the heritable quantitive traits that underlie disease[491]. GWAS is a comprehensive, hypothesis-free, and, when population stratification is controlled for, unbiased approach[491]. In statistical terms, there are broadly two types of models that are widely employed: asymptotic tests (Wald test, score test, log-likelihood ratio test (LRR)) and linear mixed effect models. All of those test are based on a classical linear regression: $Y = \alpha + \beta_1 X + \beta_2 C + \varepsilon$ for quantitative traits or $\text{logit}(Y) = \alpha + \beta_1 X + \beta_2 C + \varepsilon$ for binary traits. In the first case Y is the population mean of the trait and β is the degree of change in Y for every 1-unit of change in the predictor variable X, while in the second it is the probability of being a case [Pr(Y=1)] and β_1 is the log odds ratio for X. In both cases α is an intercept, X is the genotype/allele count, $\beta_2 C$ estimates the effect of a covariate and ε is the error term[492]. In its simplest form (without covariates), the statics for the various tests is denoted as follows:

Wald test: $W=\hat{\beta}/SE(\hat{\beta})$

LRR: $-2\,[L(\hat{\alpha},0) - L(\hat{\alpha},\hat{\beta})]$

Score test: $S=U_\beta/\sqrt{\text{var}(U_\beta)}$; $U_\beta=\partial a/\partial\beta$

These models treat the effect of the predictor variable as fixed and assume that: 1) linearity: the response variable is a linear combination of the parameters (regression coefficients) and the predictor variables; 2) homoscedasticity: the values of the response variable have the same variance in their errors, regardless of the values of

the predictor variables; 3) independence of errors: the errors of the response variables are uncorrelated with each other; 4) lack of perfect multicollinearity in the predictors: the predictors are not strongly correlated with each other. In addition, the models assume that the response variables have normally distributed errors. If that is not the case, generalised linear models where the errors are permitted to follow other distributions than the normal distribution would be preferred. For common variants (MAF > 5%), all those test give very similar results.

Finally, random effect (u) can also be introduced in the model: $Y = \alpha + \beta_1 X + \beta_2 C + u + \varepsilon$; $u \sim N(0, \sigma_g K)$ where K is a kinship matrix. In genetic epidemiology models random effects are introduced to control for population structure that is due either to true genetic structure (including family relatedness) or to pseudo structure introduced as result of sampling or genotyping procedures (using different genotyping arrays, batch effects, etc.)[492].

Using GWAS has led to the identification of 33,674 unique SNP-trait associations*, however, GWAS have certain limitations. First, the observed effects of variants are usually small and, because GWAS relies on linkage disequilibrium (LD), finding the causative variant or gene is not straightforward. The sentinel variant is more often than not a proxy for the casual variant and usually in high LD with a number of other variants[493,494]. Some of those variants may have causative effect but they may themselves be just proxies too. If a small set of likely causative variants (a.k.a credible set) is identified through fine mapping, the function of those may still be difficult to predict. This is because associated variants usually do not fall within a well-established cis-regulatory element such as a known promoter, and thus challenge predictions of their impact on regulatory elements, like disruption of transcription factor binding or function of a long non-coding RNA (lncRNA)[493,494]. Association signals tend to cluster in gene-rich areas where a number of genes are located within the same LD block, adding another level of difficulty to causative gene identification. Cell-type specific regulation may be of help here but for a lot of complex diseases the relevant cell types are not well established. Furthermore, because of chromatin looping that places regulatory enhancer elements in proximity to the promoters of genes that may be quite distant on the physical map, a causal variant may influence expression of distant genes[493,494]. The various strategies that can be used to solve those problems are discussed in more details and applied in the next chapter (**Chapter 5**), which deals with a meta-analysis of GWAS in over 19,000 people. Due to the small overall-sample size many of the loci resulting from the GWAS presented in this chapter are likely to be

* GWAS catalogue (2017-04-03): https://www.ebi.ac.uk/gwas/

false-positive, and thus, no in depth fine-mapping was attempted here but possible candidate genes were short-listed on the basis of published literature.

4.2.3. Analysis of low-frequency and rare genetic variants

Due to the small effects of variants identified by GWAS, much of the genetic variance of common traits is still unexplained[495]. One of the proposed methods to deal with this problem is the inclusion of low frequency (MAF<5%) and rarer (MAF<1%) genetic variants with moderate effects that are poorly detected by available genotyping arrays but are obtainable using sequencing[495]. Traditional regression models are not suitable for dealing with rare variants, and a variety of aggregate ("set-based", "collapsing") tests have been proposed as an alternative[496-506]. Those tests can be divided into non-adaptive and adaptive burden tests, variance component tests, and combined and exponentially combined tests[507]. Burden tests, such as the cohort allelic sums test (CAST) and the combined and multivariate collapsing (CMC) method, collapse information for multiple genetic variants into a single genetic score and test for association between the genetic score and a trait of interest[496,497,503]. CAST assumes the presence of any rare genetic variant increases the risk, while CMC categorises variants within MAF categories and gives different weights for each category[496,497]. Alternative methods employing continuous weight functions are also available[502,508]. Burden tests are powerful when a trait is explained by a large proportion of causal variants with the same direction of effect but lose power in the presence of both trait-increasing and trait-decreasing variants or a small fraction of causal variants. The adaptive burden tests are similar to the burden tests but are robust in the presence of null variants and allow for variants with different direction of effects. This is done by estimating the effect of each variant in advance and taking the direction of the effect into account (e.g. aSum)[509] in the burden calculations, by using the betas as weights (e.g. EREC)[510], or by using permutation-based frequency threshold (e.g. VT)[499] or kernel-based adaptive weighting (KBAC)[511]. The next class of tests, the variance-component tests, evaluate the distribution of genetic effects (weighted or unweighted aggregated score test statistics) for a group of variants within a random-effects model (C-alpha, SKAT, SSU)[498,500-502]. Of the three tests SKAT (SNP-set (Sequence) Kernel Association Test) is the most commonly used. This test affords the use of covariates (including kinship matrix), flexible choice of weights and a resampling procedure for empirical p-value calculation[500,502]. Finally, the burden and variance-component tests can be combined to produce a single p-value[504-506]. This p-value is obtained either by

taking the difference between the p-value from the two tests or by combining them in a linear or an exponential manner[504-506]. During my master-degree thesis project and one of my first-year PhD rotations, I explored the utility of different tests in the context of common and rare conditions* and found the different tests tend to give very different results from each other, and very inflated p-values. Of all the tests, the use of SKAT resulted in the lowest amount of inflation and in this chapter I will present the results I obtained using this test.

4.3. Materials and Methods

Detailed cohort description of the TwinsUK and the Rotterdam Study cohorts, together with detailed description of data collection, genotyping and genotype imputation procedures, are presented in **Chapter 3**. Here I will focus on phenotyping, QC procedures and describing the various statistical analyses used to obtain results in the present chapter.

4.3.1. Analysis overview

The analysis presented in this chapter can be divided into three parts where Part I is heritability analyses, Part II is the GWAS and part III is gene-based analysis of low-frequency and rare genetic variants. Part I and II are interconnected in terms of datasets used. In addition, in the Rotterdam Study, only GWAS were performed. An overview of the analysis is presented in **Figure 4.1**.

4.3.2. Phenotyping

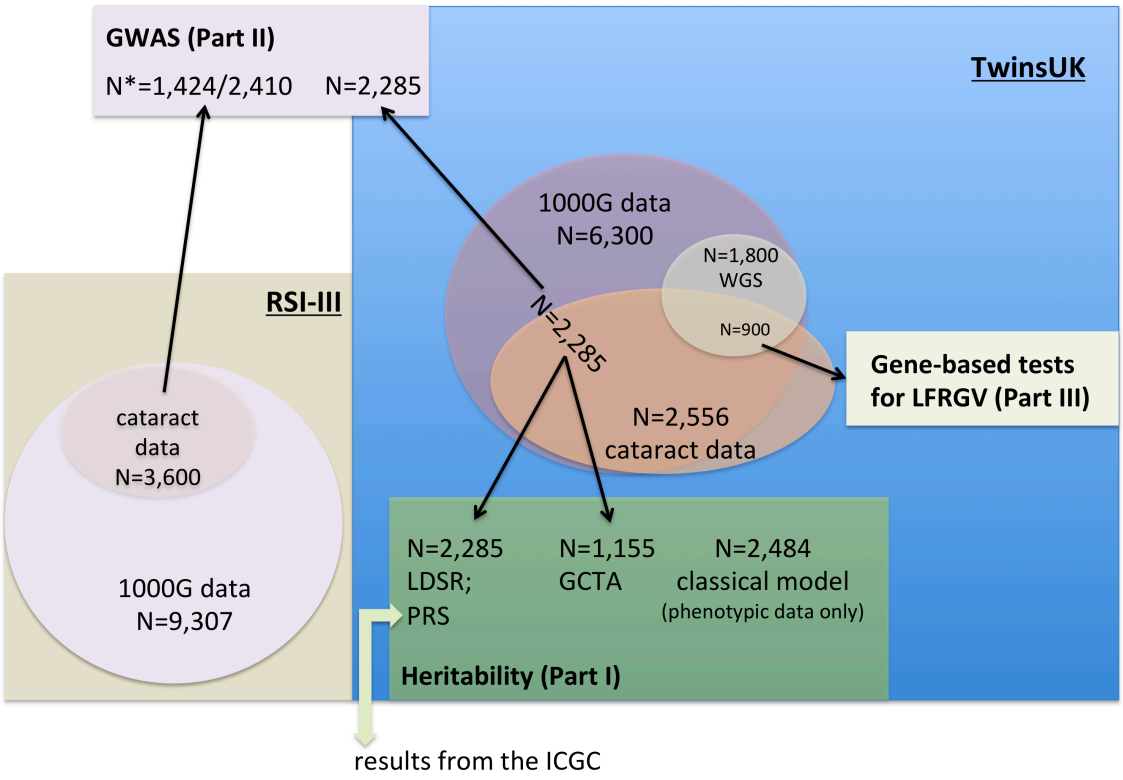
The TwinsUK cataract data was provided to me by the members of the eye-research team. My contribution to the data collection for the Rotterdam study is described in **Chapter 3** (section 3.3.2. Data collection in the Rotterdam study). I also performed the cataract grading for this cohort as described below.

TwinsUK: All subjects were unaware of eye research interests at the time of enrolment. Cataracts were assessed in all individuals over the age of fifty, as a part of a complete ophthalmological examination during the following periods (January 1998 – July 1999, May 2006 – September 2010). Digital black-and-white Scheimpflug and

* age-related cataract (subset of the TwinsUK sequencing data); Silver-Russell syndrome

Retroillumination lens photographs of dilated eyes were taken in dark conditions using

Figure 4.1.Analysis overview



This figure presents an overview of the analysis presented in Chapter 4. N — number of people; N* — number of people after exclusion of low quality images; 1000G — data imputed to the 1000 Genomes panel; WGS — whole genome sequencing data; LDSR — linkage disequilibrium score regression; GCTA — Genome-wide Complex Trait Analysis; PRS — polygenic risk score; LFRGV — low frequency and rare genetic variants; GWAS — genome-wide association analysis; ICGC — international cataracts genetics consortium; RSI-III — Rotterdam study I, follow-up III.

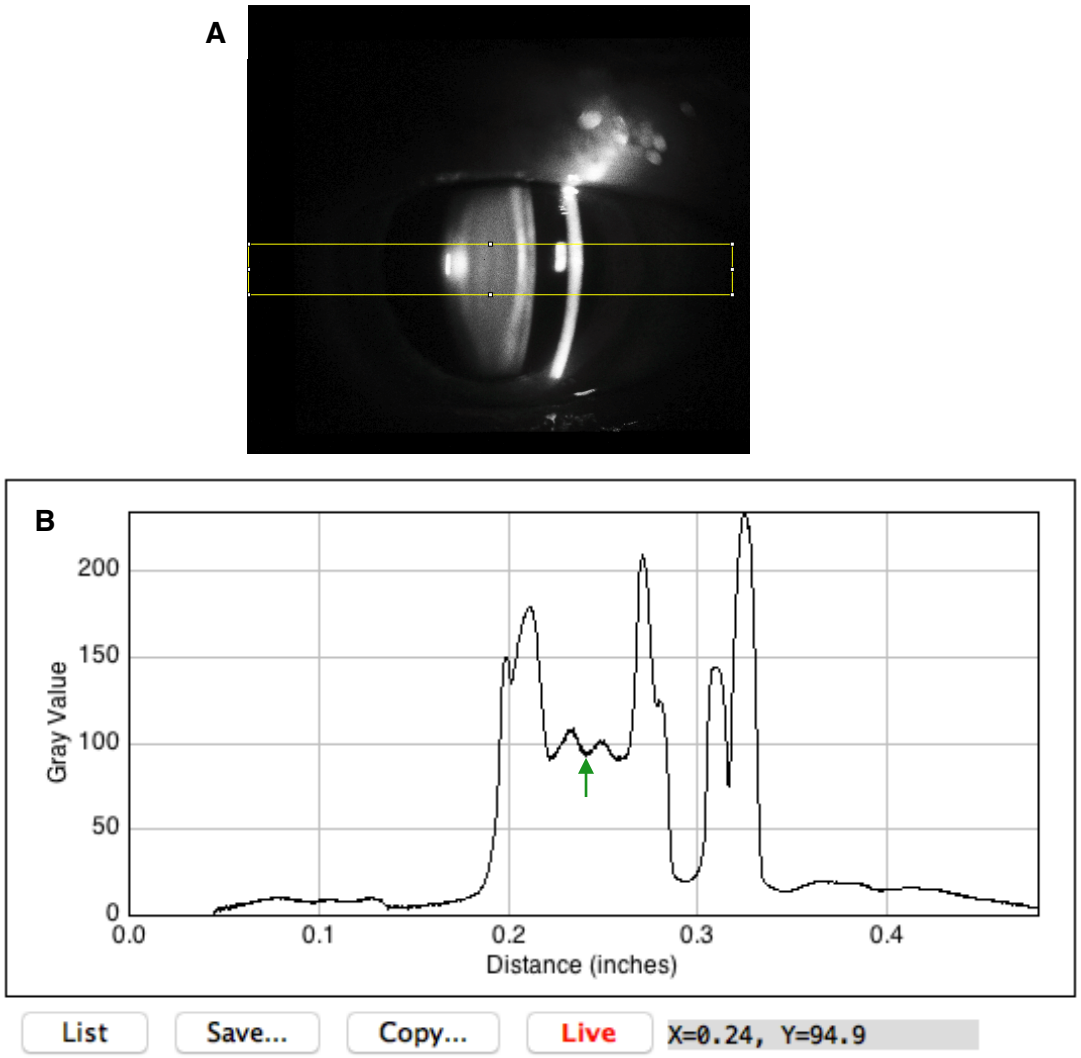
the Case2000 camera (Marcher Instruments Ltd, Worcester, UK). Only the right eye in each person was assessed, except if that eye was aphakic/pseudophakic. Cataracts were not assessed in participants who had bilateral cataract surgery or were uncomfortable with being dilated. The total number of TwinsUK participants with gradable cataract photos was 2,556.

Nuclear cataract was graded using an objective, densitometry-based grading system that measures the pixel density at the centre of the lens nucleus (central nuclear dip score (CNDS))[167,512]. This score (range: 0-255) measures the amount of white scatter (or opalescence) in the lens, but not of brunescence due to images being

in grey scale. In TwinsUK the nuclear cataract grading was done automatically by software inbuilt in the camera. Subsets of TwinsUK images were also graded using the Modified Oxford grading system[169] and a CNDS of ≥ 70 wad previously calculated as equivalent to clinically significant cataract according to that system.

Cortical cataract was graded subjectively using modified Modified Oxford grading system[169].

Figure 4.2.Nuclear cataract grading in RSI-III



This figure shows a snapshot of the nuclear cataract grading set up: A) a lens image with the operator-defined measurement window in yellow; B) the density function read out from A) with a green arrow pointing to the the central nuclear dip measurement point. The coordinates of that point are shaded in grey; In this case the central nuclear dip score is 94.9.

Rotterdam Study (RSI-III): All subjects were unaware of eye research interests

at the time of enrolment. As part of the third follow-up of the Rotterdam Study (January 1998 to December 1999), 3,600 participants attended an eye test. During the eye test lens images were taken (N=15,599) for both eyes using an analogue camera. Nuclear cataracts were assessed using Slit Lamp/Scheimpflug (1 image per eye), while cortical and PSC cataracts (2 images per eye) were assessed using Neitz CT-R camera (Kowa Optical, Torrance, CA).

The grading of nuclear cataract was designed to closely follow that of the TwinsUK cohort, but was semi-automated. Each image was loaded separately in ImageJ (https://imagej.nih.gov/ij/) and a macro designed by Dr. Zoë Forkin was launched. The macro plots the grey pixel density function across the lens in an operator-specified measurement window. The operator then locates the central nuclear dip and registers the density value at its lowest point in a separate excel spreadsheet (**Figure 4.2.**). Nuclear cataract in this cohort was not graded using any other system, thus it could not be determined which grades were equivalent to clinically significant cataract.

Similarly to TwinsUK, cortical cataract for the RSI-III cohort was also graded using the Oxford grading system. Clinically significant cortical cataract, according to that system, is defined as ≥ 5% of lens area affected by cortical spokes.

4.3.3. Statistical analysis

Heritability analysis: Classical heritability analyses were carried out in 1232 twin pairs (620MZs and 612DZs) using univariate models and bivariate Cholesky decomposition modelling implemented in OpenMx (https://openmx.ssri.psu.edu/). Two different models were explored. In the first model, both nuclear and cortical cataract were treated as quantitative traits but the cortical cataract scores were not normalised. As the heritability modelling assumes that variables are normally distributed, a second model where the quantile normalised cortical cataract scores were used was explored. In this case, to make the nuclear scores comparable to the cortical scores, the nuclear scores were adjusted for age and sex, and used the residuals.

Two different methods, GCTA and LD-score regression, were used to calculate SNP-based bivariate heritability. In order to make the GCTA and LD-score regression analyses comparable, the GCTA relatedness matrix was calculated using the SNPs implemented in the LD-regression score program (N=1,027,466, high quality HapMap3 SNPs). In addition, as GCTA is unable to handle twin data, bivariate heritability analysis were carried in 1155 individuals (one twin per pair) and nuclear and cortical cataract

were treated either as both quantitative or as both case-control traits, in the latter case the disease prevalences detected in TwinsUK were used: 22% for nuclear and 16% for cortical cataract.

Polygenic risk scores: For each SNP, the reference allele, beta coefficient and p-value from the GWAS meta-analysis of nuclear cataract, excluding the TwinsUK cohort were extracted. Detailed information about this meta-analysis is presented in **Chapter 5**. The extracted SNPs were then pruned for linkage disequilibrium using PLINK (pairwise cut-off of $r^2 \leq 0.25$ within 200 kB window, and a MAF cut-off of 5%). Polygenic risk scores for nuclear cataract were calculated with PLINK at the following p-value thresholds: $< 0.0001, < 0.001, < 0.01, < 0.05, < 0.1, < 0.5$. The effect of the risk score on cortical cataract was calculated by using a linear model adjusting for age and sex. Risk scores with uncorrected p-value <0.05 for association were considered statistically significant. Variance explained (r^2) was calculated using logistic regression with and without covariates. In all cases family structure was taken into account.

As the results from the cortical cataract meta-analysis were not available to me, I could not calculate the effect of cortical cataract risk score on nuclear cataract.

Genome-wide association analysis: The cataract scores for nuclear and cortical cataracts in the two cohorts were not normally distributed and, given that the GWAS linear models work optimally for normally distributed data, all cataract scores were transformed. The nuclear score was normalised using natural logarithm. In addition, to permit meta-analysis at a later stage (**Chapter 5**), the transformed scores were standardised (subtracting the mean and dividing by the standard deviation). Due to the high percentage of people having a score of zero, the cortical cataract scores are not easily transformable. Cortical cataract score were therefore treated as binary trait with a case-control status cut-off of 0.25 which is equal to 5% of lens area affected by cortical spokes.

In GWAS, confounding due to population structure causes inflation of the test statistics and high-false positive rates. In homogeneous cohorts, such as TwinsUK and RSI-III (**Figure 4.2.**), excluding ancestry outliers as described above is usually enough to prevent inflation. In more heterogeneous cohorts, adjusting for genetic principal components is recommended. The TwinsUK cohort consists of related individuals which creates structure in the data that needs to be accounted for. Therefore, in TwinsUK, GWAS analyses were performed using a linear mixed effect model implemented in GEMMA[513,514] that includes a relatedness matrix* within the random

* the matrix was calculated using standardised genotypes to account for the fact that lower minor allele frequency tend to have larger effects

$$\mathbf{Y} = \mathbf{WA} + \mathbf{x}\boldsymbol{\beta}^T + \mathbf{U} + \mathbf{E}; \quad \mathbf{G} \sim \mathrm{MN}_{n\times d}(\mathbf{0}, \mathbf{K}, \mathbf{V}_g), \quad \mathbf{E} \sim \mathrm{MN}_{n\times d}(\mathbf{0}, \mathbf{I}_{n\times n}, \mathbf{V}_e),$$

effect term:

where Y is a vector of phenotype values; W is a matrix of fixed effect covariates; A a matrix of the corresponding coefficients for this covariates, including intercept; x is a vector of SNP genotypes (or allele dosage probabilities in the case of inputed variants); β is a vector of SNP effect sizes; U is a matrix of random effects; E is a matrix of errors; K is a relatedness matrix, I is identity matrix, Vg is a symmetric matrix of genetic variance components; and Ve is a symmetric matrix of environmental components. Both the genetic and environmental effects are normally distributed.

In RSI-III, GWAS was performed using a standard linear and logistic regression models implemented in MACH2QTL and MACH2DAT[515]:

$$\mathbf{y = WA + \beta X + E}$$

$$\mathbf{odds(y) = exp(WA + \beta X + E)}$$

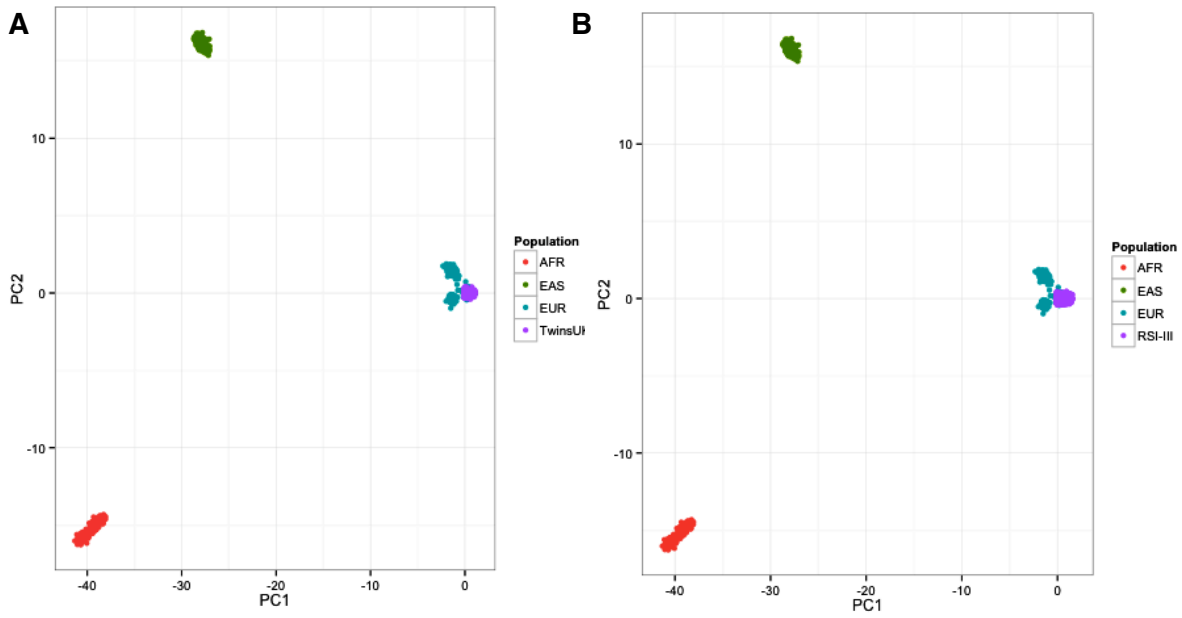
where Y is a vector of phenotype values; W is a matrix of fixed effect covariates; A a matrix of the corresponding coefficients for this covariates, including intercept; X is a vector of allele dosage posterior probabilities for each SNP; β is a vector of SNP effect sizes; and E is a matrix of errors.

All models included age and sex as covariates and in the case of nuclear cataract, a model using smoking status as a covariate was also explored. The percentage of non-, ex- and current smokers in TwinsUK were: 55.8, 29.6, and 14.6 respectively. The percentage of non-, ex- and current smokers in RSI-III were: 44.2, 2.5, and 53.3 respectively.

GWAS results QC: I performed QC of the summary statistics from the separate GWAS using a protocol developed by the Genetic Investigation of Anthropometric Traits (GIANT) Consortium[516]. First, I confirmed that variants aligned to the positive stand of build GRCh37/hg19 and that the frequencies of both the imputed and genotype variants did not deviate significantly from those reported by the 1000G (Pearson correlation coefficient larger than 0.7). Then, variants were excluded if they: had low imputation scores (< 0.30); deviated from HWE (p-value < $1x10^{-4}$); effect sizes that were outliers in from the distribution of effects (below the 0.1 and the above the 99.9 percentiles); SNPs with sample size less than 30 and minor allele count less than 6. In addition to these exclusions, as the 2012 version of 1000G is less reliable for

insertion-deletion polymorphism, those variants were excluded from the RSI-III analysis.

Figure 4.3. Principale component plots for TwinsUK and RSI-III



This figure shows the results of the principale component analysis for TwinsUK (A) and RSI-III (B) together with the 1000 Genomes project populations. For the purposes of a clearer presentation the South Asian, Amerindians and Afro-Americans have been excluded from the plot. Both TwinsUK and RSI-III cluster with the European ancestry populations (on top of the British and Utah Residents with Northern and Western European Ancestry and slightly apart from the Finnish and Tuscany clusters). PC – principale component; AFR – African ancestry populations; EAS – East Asian Ancestry populations; EUR – European ancestry populations.

Analysis of low-frequency and rare genetic variants (LFRV): Of the 1800 individuals with genome-wide sequencing data, 905 (100% female) had cataract phenotypes. Similarly to the GWAS analysis, nuclear cataract was treated as quantitative trait and cortical cataract was treated as binary trait, and all analyses were adjusted for age. Low frequency and rare variants are much more susceptible to subtler effects of population substructure than common variants[517-519]. This can partially explain the higher inflation factors seen when such variants are analysed. The other reason being that LFRV effects are more likely to be sampled from the tails of the effects distribution. Thus, even though the 1800 individuals are unrelated and ethnically white British, a kinship matrix was used in the analysis. These resulted in substantially reducing the inflation of the test statistics. Exome-centric aggregate analyses were

conducted for both types of cataract using SNP-set (Sequence) Kernel Association Test (SKAT) implemented in EPACTS software*:

$$Q = \left(\mathbf{y} - \widehat{\boldsymbol{\mu}}\right)' \mathbf{K} \left(\mathbf{y} - \widehat{\boldsymbol{\mu}}\right)$$

where Q is a variance-component score statistic that equals a weighted linear kernel function times the difference between the phenotypic value of an individual and the predicted mean value for the phenotype as function of the covariates[502]:

$$\widehat{\boldsymbol{\mu}} = \hat{\alpha}_0 + \mathbf{X}\widehat{\boldsymbol{\alpha}} \qquad \text{quantitative traits}$$

$$\widehat{\boldsymbol{\mu}} = \text{logit}^{-1}\left(\hat{\alpha}_0 + \mathbf{X}\widehat{\boldsymbol{\alpha}}\right) \qquad \text{binary traits}$$

The kernel function (K = GWG') measures the genetic the similarity between individuals. The weights (W) per variant are drown from a beta distribution function

$$\sqrt{w_j} = Beta\left(MAF_j;\ a_1, a_2\right)$$

where a_1 and a_2 are pre-specified parameters evaluated at the sample MAF for the j-th variant in the data[502].

Synonymous changes and sets where the signal came from less than 3 variants were excluded. Genes known to consistently generate false positive signals were also excluded[520]. All variants were annotated against the human genome build GRCh37/ hg19 and NCBI human RefSeqGenes (19,821). All genes with p-value < 2.52×10^{-6} were considered significantly associated with cataract after Bonferroni correction for multiple testing (0.05/19,821).

Result visualisation: Plots were created using STATA14 and ggplot2 under R (https://www.r-project.org/). Regional association plots were created using LocusZoom (http://locuszoom.org/) as follows: the hg19/1000 Genomes (November 2014) European reference panel was used as an LD-reference panel for TwinsUK, while hg19/1000 Genomes (March 2012) was used an LD-reference panel for RSI-III.

4.4. Results

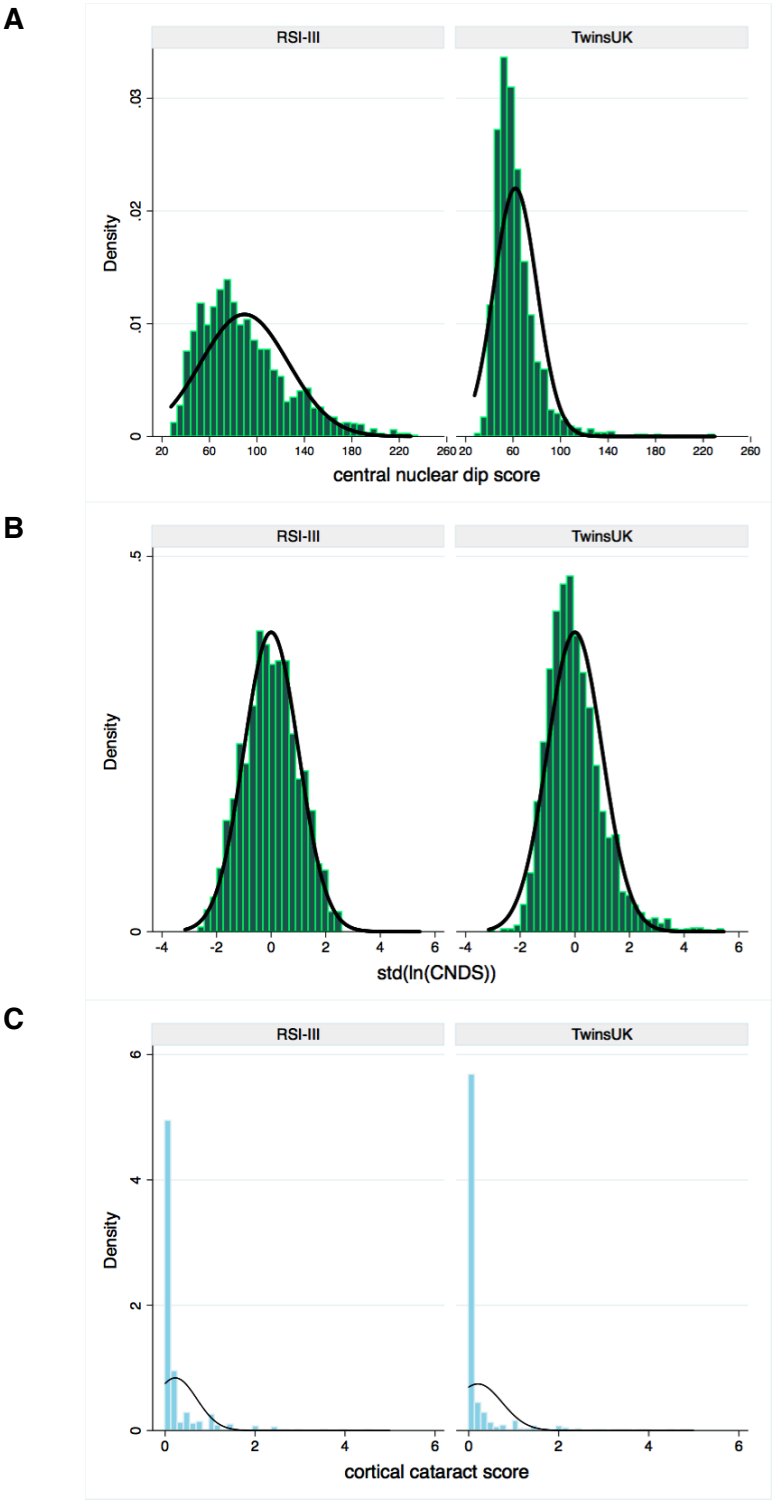
In TwinsUK, cataract was measured in 2556 predominantly female (98%) twins, with mean age at eye visit of 62 years (range: 50-83 years). Monozygotic (MZ) and dizygotic twin (DZ) were equally represented (MZ:DZ ratio = 0.8). In RSI-III lens photographs† were available for 2,773 participants, who were also predominantly female (66%) but, in contrast to the TwinsUK sample, were on average 10 years older (mean age at eye visit of 71.9 [61-100 years]). After excluding RSI-III individuals with

* http://genome.sph.umich.edu/wiki/EPACTS

† aphakic and pseudophakic eyes

bilateral cataract surgery and with poor quality photographs, nuclear and cortical cataracts were graded in 1,589 and 2,434 individuals respectively. The excluded RSI-III individuals were on average older (4.6 years, p-value$_{t\text{-}test}$ < 0.0001) than the included

Figure 4.4: Distribution of cataract scores in TwinsUK and RSI-III cohorts



This figure shows the distribution of nuclear cataract before (A) after (B) transformation, and the distribution of cortical cataracts (C) in TwinsUK and RSI-III. In TwinsUK only one eye per person was available, while in RSI-III the eye with the higher cataract score was used. CNDS - central nuclear dip score; std - standardised; ln- natural logarithm.

individuals but did not differ from the latter in terms of sex or smoking status ($p\text{-value}_{z\text{-test}} > 0.05$). Cataract scores summary statistics (**Table 4.1.**) and score distributions (**Figure 4.4.**) for both cohorts are presented below. Both nuclear and cortical cataract scores were significantly higher (p-value < 0.001 from Mann-Whitney rank-sum and χ^2 tests respectively) in RSI-III than in TwinsUK. Theses differences may be due to the fact that RSI-III individuals were older but may also be due to differences in cameras, graders, etc.

Table 4.1: Summary statistics for nuclear and cortical cataract scores

	Nuclear cataract scores			Cortical cataract scores				
cohort	mean (SD)	range	CS (%)	0 (%)	> 0			CS (%)
					median	range	IQR	
TwinsUK	62.1 (18.1)	34.5-229.2	22.0	57.9	0.15	0.02-5.0	0.4	16.3
RSI-III	89.6 (36.9)	30.3-229.8	NA*	54.6	0.25	0.01-4.5	0.5	25.6

This table summarises the nuclear and cortical cataract scores for the TwinsUK and RSI-III cohorts. Nuclear cataract: CS - % of people with clinically significant cataract as compared against the Modified Oxford Grading System. * Nuclear cataract in the Rotterdam study was graded using CNDS only, thus it could not be determined which CNDS cut-off in this study was equivalent to a clinically significant cataract as measured by a clinically relevant grading systhem. Cortical cataract: the % of people with score of 0 is reported as well as summary statistics for individuals with non-zero scores. IQR - interquartile range.

4.4.1. Heritability analysis

Before focusing on bivariate heritability modelling, I first calculated the univariate heritability for nuclear and cortical cataract (**Table 4.2**). In the adjusted model, the heritability of nuclear cataract was 0.64 [95%CI: 0.60-0.68] and that of cortical cataract was 0.65 [95%CI: 0.61-0.69]. The rest of the variance for both nuclear and cortical cataract was explained by environmental factors: 0.36 [95%CI: 0.32-0.40] and 0.35 [95%CI: 0.31-0.39] respectively.

These heritability estimates were higher than the ones previously obtained[167,169].

The observed difference could be explained as follows. First, the heritability in the present thesis was calculated in larger sample (620MZ and 612DZ) which affords an increase in power compared to the Hammond *et al*. studies[167,169]. This is also exemplified by the tighter confidence intervals obtained here. Second, when the effects of strong environmental contributors are removed from a heritability model, the heritability estimates usually increase. In terms of nuclear cataract, this means comparing heritability calculated from raw cataract scores[167] with the heritability calculated from the residuals of a linear regression with log-transformed CNDS as a dependent variable and age and sex as predictors. While for cortical cataract the comparison is between a liability model with age as a latent variable[169] and a linear regression model with age and sex as covariates. Last but not least, using normally distributed residuals in modelling affords more accurate estimates than using row variables.

Table 4.2: Univariate heritability analysis not nuclear and cortical cataract

ICC MZ	ICC DZ	Model	-2 LL	df	AIC	diffLL	diffdf	p-value
				Nuclear cataract				
0.65	0.36	ACE	19797.09	2440	14917.09	-	-	-
		AE	19799.58	2441	14917.58	2.49	1	1.1E-01
		CE	19843.95	2442	14961.95	46.87	1	7.6E-12
		E	20215.91	2442	15331.91	418.82	2	1.1E-91
				Cortical cataract				
0.66	0.41	ACE	6483.98	2440	1603.98	-	-	-
		AE	6491.52	2441	1609.52	7.54	1	6.0E-03
		CE	6522.66	2442	1640.66	38.68	1	5.0E-10
		E	6939.80	2442	2055.80	455.82	2	1.1E-99

Model-fitting results for univariate analysis of nuclear and cortical cataract against the full (ACE) model. Intraclass correlation coefficient (ICC); -2 log-likelihood (-2LL); df – degrees of freedom; AIC — Akaike information criterion used to evaluate the most parsimonious model; diffLL – difference between the log-likelihood estimates of the models; diffdf – difference in number of degrees of freedom between models; A — additive genetic component, C — common environment, E unique environment and measurement error.

Next, I performed classic bivariate heritability analysis (**Table 4.3**). In this sample the two traits were weakly correlated (r_S= 0.13). Additive genetic factors explained 34% of variance in the unadjusted nuclear score and 56% of variance in the unadjusted

cortical cataract scores, with the remaining variance being explained by unique environmental factors (AIC = 16523.54) (**Table 4.3.**). Similarly to the correlation above, in this model nuclear and cortical cataract scores shared 13% of phenotypic variance and 32% of this shared variance could be attributed to genetic factors (p-value = 0.03). Using the age- and sex-adjusted quantitative traits significantly changed the results and yielded a larger heritability estimate (**Table 4.3.**). The confidence intervals were, however, much larger and the model did not have a better fit (AIC = 16523.00). In the adjusted model the two traits shared only 3% of phenotypic variance and 76% of that was due to genetic factors (p-value = 0.54). When applying a bivariate model, the point estimates yielded in both cases were mostly lower than those from the univariate model (0.64 [0.60-0.68]). Such drop in the effect size is frequently observed and may happen because adding a variable changes the amount of information about the genetic variance component. Alternatively, the drop may also result from measurement precision differences between the two variables in the model. The latter is far more likely in the case of the analyses presented here.

Table 4.3: Comparison between bivariate heritability estimates obtained using three different strategies

analysis type	V_A [95%CI] or (SE)		R_{ph} [95%CI]	C(G)	h^2 [95%CI]
	nuclear cataract	cortical cataract		shared	
Cholesky decomposition (N=2,464)					
unadjusted	34% [23%-45%]	56% [49%-62%]	13% [8%-18%]	NA$	32% [10%-94%]
adjusted	44% [0.3%-59%]	51% [0.4%-66%]	3% [-40%-12%]	NA$	76% [19%-89%]
GCTA (N=1,155)*					
both quantitative	28.4% (8.7%)	9.0% (6.8%)	NA$	8.7% (1.7%)	54.7% (12.8%)
both binary	6.9% (3.4%)	5% (4.7%)	NA$	0.5% (2.8%)	9.3% (47.8%)
LDSR (N=2,285)					
adjusted**	9.0% (7.2)	5.2% (2.5%)	NA$	0.1% (2.8%)	NA***

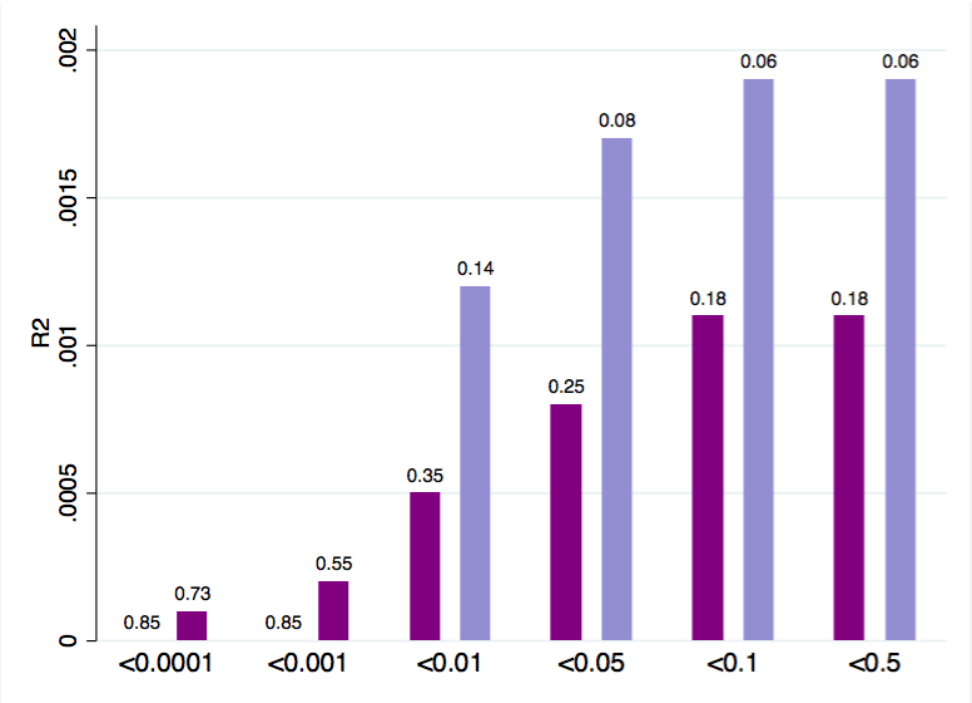
Comparison of bivariate heritability estimates between bivariate Cholesky decomposition, Genome-wide Complex Trait Analysis (GCTA) and linkage disequilibrium score regression (LDSR) models. V_A – % of variance explained by additive genetic factors (h^2 in the classical model and h^2_{SNP} in the other two models); CI – confidence interval; SE – standard error; R_{ph} – amount of shared phenotypic variance; C(G) – genetic covariance between 2 traits; h^2 – amount of shared phenotypic variance explained by genetic factors (h^2 in the classical model and h^2_{SNP} in the other two models); N – number of people used for the analysis; unadjusted – raw scores; adjusted – analysis adjusted for age and sex; * adjusted for age and sex; NA$ – not outputted by the program; ** calculated using GWAS summary statistics where GWAS analysis were

adjusted for the effects of age and sex and for relatedness; *** due to very low genetic covariance, the amount of shared heritability could not be calculated.

When both traits were treated as quantitative using GCTA (N=1,155), the proportion of variance explained by SNPs for nuclear and cortical cataract were 28.4% and 9.0% respectively. In this model the genetic covariance between the two traits was 8.7% and genetic correlation was 54.7%. When both traits were treated as binary, a less powerful model, genetic variance for nuclear and cortical cataract were 6.9% and 5.0% respectively, while the genetic covariance and the genetic correlation between the two traits were 0.5% and 9.3% respectively. The LD-score regression analysis (N = 2,285) estimated the heritability of nuclear and cortical cataract to be 9.0% and 5.2% respectively. The genetic covariance between the two traits was very low (0.1%) and, thus, correlation could not be calculated.

Finally, a polygenic risk score for nuclear cataract, calculated using the meta-analysis results from the International Cataract Genetics Consortium (TwinsUK cohort excluded), could predict only a very small amount of variance in cortical cataract in TwinsUK (**Figure 4.5.**)

Figure 4.5. Genome-wide polygenic scores prediction



This figure presents the results of the amount of variance in cortical cataract explained by nuclear cataract associated variants. Univariate (purple) and adjusting for age and sex (lavender) analysis. The number of SNPs per p-value cut-off are (from left to right): 204, 1,768,

16,184, 78,120, 153,800 and 747,352.

4.4.2. Genome-wide association analysis results

In TwinsUK of the 2556 people who had cataract scores, 2285* people had genome-wide data imputed against the 1000G project. Of the 9.1 million variants (including indels), 9 million variants passed QC (**Figures 4.6**). The genomic inflation factor of the summary statistics was (λ) 1.02 and 1.01 for nuclear and cortical cataract respectively, indicating good control of population stratification. In the case of nuclear cataract (**Table 4.4.**), only one variant (rs79717917, MAF=2%, intronic, *TDRD3*), reached genome-wide significance (p-value < 5.8×10^{-8}, **Figure 4.7**). In addition, 153 variants at 47 loci were suggestively associated (p-value < 1×10^{-5}) with nuclear cataract. These variants were within or in close proximity to 40 protein coding genes, 3 miRNAs and one lncRNA. Adding smoking status as a covariate did not change the results for the existing associations and did not lead to the identification of any additional variants.

In the case of cortical cataract, two variants, rs116848423 (MAF=1%, *OR8B8*) and rs145013270 (MAF=1%, *C14ofr177*), reached genome-wide significance (p-value < 5.8×10^{-8}) and 121 variants at 42 loci were suggestively associated (p-value < 1×10^{-5}) with cortical cataract. The suggestively associated variants were within or in close proximity to 35 protein-coding and 2 lncRNA-coding genes. In TwinsUK, only variants within one gene (*SOX5*) were associated with both types of cataract. However, the variants associated with nuclear cataract were situated within the gene body, and those associated with cortical cataract were situated within the gene promotor (**Figure 4.8.**).

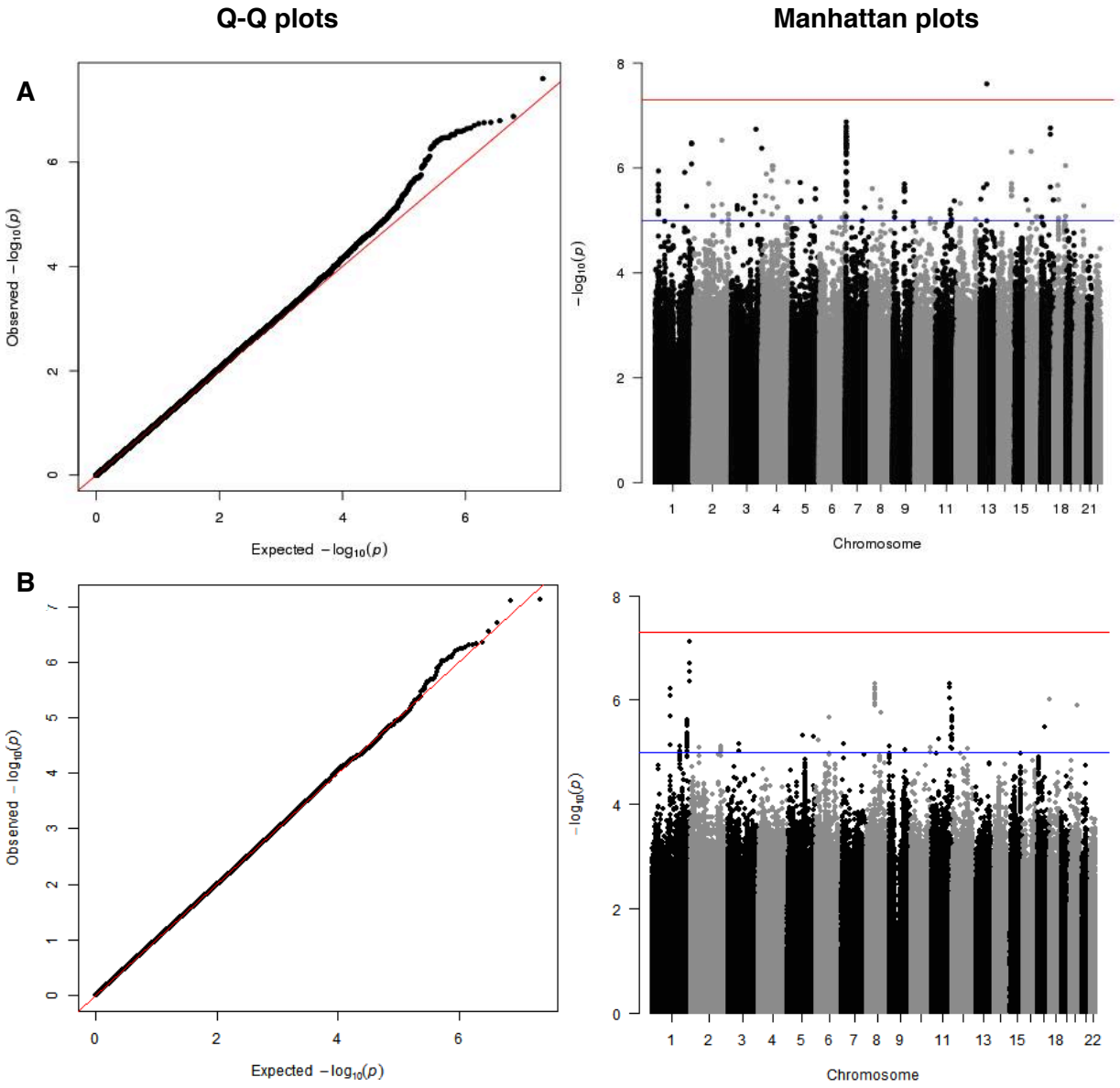
In RSI-III the overlap between genome-wide data imputed against the 1000G project (28.7 million variants, excluding indels) and nuclear and cortical cataract was 1424 and 2410 people respectively. After QC, 10 million variants were associated with both types of cataract. The λ equalled 0.99 and 1.02 for nuclear and cortical cataract respectively, indicating good control of population stratification. None of the nuclear cataract associated variants reached genome-wide significance (**Figure 4.6, Table 4.4.**). In addition, 22 variants at 5 loci were suggestively associated (p-value < 1×10^{-5}) with nuclear cataract. These variants were within or in close proximity to 5 protein coding genes. Adding smoking status as a covariate did no change the results for the

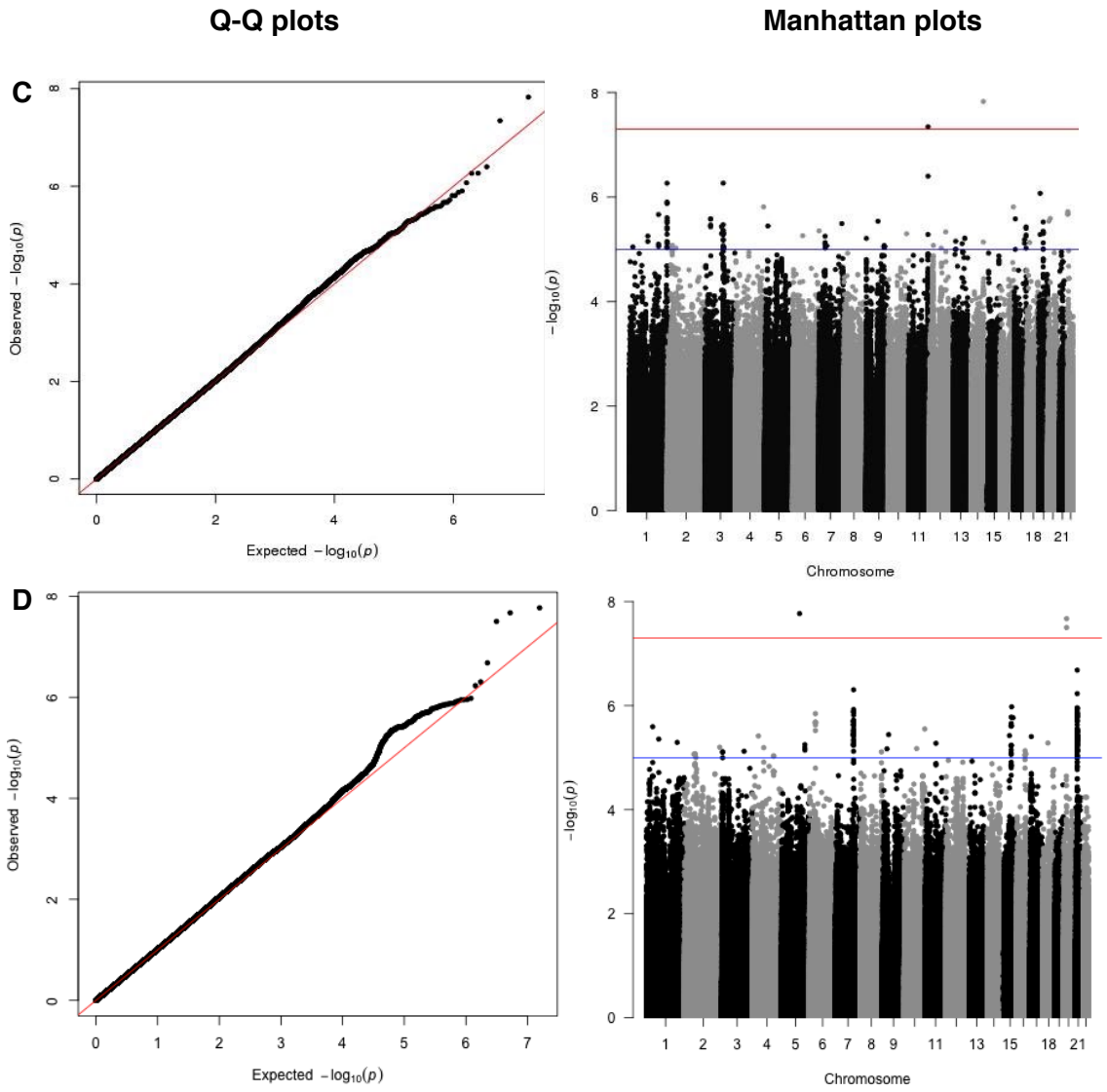
*for the majority of twin pairs one twin per pair at random was included in the imputation; while the genotypes of MZ twins could be copied from one twin to the other, that was not possible for DZ twins.

existing associations and did not lead to the identification of any additional variants.

As opposed to nuclear cataract, three variants within 2 loci, *LINC02214* and *C20orf78*, reached genome-wide significance (p-value < 5.8×10^{-8}), and 193 variants at 30 loci were suggestively associated with cortical cataract (**Table 4.4**). Furthermore, in RSI-III no loci were associated with both nuclear and cortical cataract, and the association with *SOX5* did not replicate (p-value > 0.05). *SOX5* is plays important role in various developmental processes but its role in lens development is unclear[521].

Figure 4.6. Quantile-quantile (Q-Q) and Manhattan plots of the GWAS results for nuclear and cortical cataracts





This figure presents the GWAS results for nuclear (A,B) and cortical cataracts (C,D) in the TwinsUK (A,C) and RSI-III cohorts (B,D). Q-Q plots: the observed p-values distribution is plotted agains the expected p-values distribution under the null hypothesis. Manhattan plots: p-values are plotted against the 22 autosomes. red line – genome-wide significance threshold (p-value=$5x10^{-8}$); blue line – suggestive significance threshold (p-value=$1x10^{-5}$). All p-values are -log10 transformed.

For both traits, over 50% of associated variants in TwinsUK were of low MAF (between 1% and 5%) and could not be replicated in RSI-III because the SNPs were either insertion-deletions (not considered for RSI-III, see methods), were of low imputation quality (< 0.3) or were absent in RSI-III. The most strongly associated loci were different in the two cohorts and did not replicate each other in terms of p-value

(**Table 4.4.**) but for 41% of sentinel variants the direction effect was similar for both cohorts. The rest had opposite effect in the two cohorts but for those variants there was low statistical confidence in the direction of effect. For example, in the cases where there was no statistically significant association in the RSI-III, the effect estimates mostly had very wide confidence intervals passing through zero. In the cases where was no statistically significant association in the TwinsUK, the confidence intervals were tight but the effect was centred on zero. Moreover, as illustrated by **Figure 4.7.**, none of the SNPs in LD with the sentinel variant showed association in the replicating cohort. Finally, similarly to the shared heritability results, there was no overlap between the variants or loci that were associated with nuclear cataract and those that were associated with cortical cataract (**Table 4.4.**).

As the following chapter (**Chapter 5**) is dedicated to meta-analysis of cataract GWAS, I did not attempt to meta-analyse the two cohorts in the present chapter. However, the sentinel variants were meta-analysed using fixed effect inverse-variance meta-analysis implemented in METAL[522] (**Table 4.4**). These revealed substantial heterogeneity of effects between the two cohorts for the majority of sentinel variants and resulted in higher meta-analysis p-values (**Table 4.4.**). The only exceptions were two of the variants that were associated with nuclear cataract: rs12991805 (p-value=7.5×10^{-9}; within *LINC01934*) and rs149786171 (p-value=1.2×10^{-7}; within *PRR11*). The meta-analyses for these variants produced lower p-values and there was no heterogeneity of effects detected in either case. These two loci will be discussed in more detail later in this chapter. Here it is worth noting that heterogeneity estimates tend to be unstable when the number of cohorts meta-analysed is small and can not be used to claim general discordance between TwinsUK and RSI-III. In fact for the majority of variants showing at least nominal significance (p-value < 0.05) the direction and effect sizes for the two cohorts were comparable.

Table 4.4: Loci associated with nuclear and cortical cataracts

position		SNP	Gene	EA/	TwinsUK cohort					RSI-III cohort					meta-analysis			
CHR	BP			OA	EAF	Beta	SE	p-value	INFO	EAF	Beta	SE	p-value	INFO	Beta	SE	p-value	$Hetl^2$
Nuclear cataract																		
1q25.3	183542747	rs3818364	*NCF2* [b]	T/G	0.49	0.04	0.03	0.14	0.99	0.45	0.17	0.04	7.9E-06	0.99	0.09	0.02	2.9E-04	85.2
1q42.3	235139145	rs75793049*	*TOMM20* [d]	G/A	0.03	0.14	0.03	3.4E-07	0.50	NA	NA	NA	NA	NA	-	-	-	-
2q31.3	182078224	rs12991805	***LINC01934*** [b]	T/G	0.01	0.17	0.03	3.0E-07	0.87	0.01	0.25	0.21	0.24	0.82	**0.17**	**0.03**	**7.5E-09**	**0.0**
2q32.3	195986610	rs2566717	*LOC105376755* [b]	C/T	0.06	0.03	0.06	0.59	0.97	0.07	-0.32	0.07	7.9E-06	0.99	-0.12	0.05	1.0E-02	93.1
7p22.3	2418116	rs6461350	*EIF3B* [b]	A/T	0.20	0.05	0.01	1.3E-07	0.95	0.19	-0.01	0.06	0.87	0.72	0.05	0.01	9.4E-07	0.0
8q12.1	56457594	rs62516394	*XKR4* [d]	A/G	0.31	-0.02	0.03	0.55	0.99	0.31	0.20	0.04	4.9E-07	0.99	0.06	0.02	1.0E-02	94.8
8q12.3	65614741	rs2884074	***CYP7B1*** [a]	T/G	0.23	0.04	0.01	4.1E-06	G	0.23	-0.03	0.04	0.95	G	0.04	0.01	2.0E-03	65.3
9q21.12	73486034	rs1337010	***TRPM3*** [b]/ ***MIR204***	A/G	0.20	0.04	0.01	2.0E-06	G	0.19	0.07	0.05	0.11	0.99	0.04	0.01	2.1E-05	0.0
10q26.13	124402495	rs77416776	*DMBT1* [b]	T/A	0.04	0.05	0.08	0.59	0.61	0.04	0.76	0.17	1.0E-05	0.31	0.18	0.07	1.0E-02	93.0
11q25	131515796	rs2174644	*NTM* [b]	C/T	0.13	-0.01	0.04	0.72	G	0.13	-0.27	0.06	1.5E-06	0.99	-0.09	0.03	1.0E-02	92.3
13q21.2	61112354	rs79717917	***TDRD3*** [b]	C/T	0.02	0.15	0.03	2.5E-08	0.70	0.03	-0.15	0.14	0.27	0.57	0.14	0.03	3.1E-06	77.2
17q22	57249489	rs149786171	***PRR11*** [b]	A/T	0.01	-0.21	0.04	1.7E-07	0.59	0.01	-0.18	0.28	0.53	0.36	**-0.21**	**0.04**	**1.2E-07**	**0.0**
Cortical cataract																		
1q44	245948059	rs12750847	SMYD3 [b]	A/G	0.09	0.11	0.02	5.4E-07	0.98	0.08	-0.04	0.06	0.49	0.98	0.10	0.02	5.5E-07	89.2
3q13.32	118173191	rs112497433	intergenic	T/G	0.01	0.29	0.06	5.4E-07	0.90	0.01	-0.18	0.14	0.15	0.84	0.22	0.06	8.3E-05	89.5
5q23.1	116194282	rs35479485	*LINC02214* [c]	A/G	0.03	0.02	0.04	0.61	0.73	0.02	2.07	0.38	1.7E-08	0.36	0.04	0.04	0.29	96.5
7q31.2	116163103	rs6466582	***CAV1*** [b]	G/A	0.16	0.002	0.02	0.90	0.99	0.17	0.46	0.09	5.0E-07	0.94	0.02	0.02	0.31	96.0
9q31.3	112753883	rs12000943	***PALM2*** [d]/ *AKAP2*	G/A	0.34	0.05	0.01	2.1E-05	G	0.32	-0.04	0.03	0.24	G	0.04	0.01	1.6E-05	87.7
11q24.2	124333054	rs116849423	*OR8B8* [c]	A/G	0.01	0.31	0.05	4.9E-09	0.88	0.01	-0.25	0.17	0.16	0.82	0.27	0.05	3.2E-08	90.0
14q.32.2	99374266	rs145013270	*C14orf117* [d]	G/A	0.01	0.33	0.06	1.5E-08	0.89	0.01	0.06	0.15	0.67	0.90	0.29	0.06	1.5E-07	64.2
15q23	71850678	rs62022770	*THSD4* [b]	C/T	0.11	-0.03	0.02	0.09	0.90	0.10	-0.58	0.12	1.1E-06	0.86	-0.05	0.02	2.0E-02	95.1
17q25.3	76897347	rs4789863	***TIMP2*** [b]	T/C	0.04	0.13	0.03	3.7E-06	G	0.04	0.01	0.08	0.90	0.99	0.11	0.03	4.1E-05	49.3
19p13.2	8829104	rs55844979	*OR2Z1* [c]	A/G	0.06	0.13	0.03	8.5E-07	0.79	0.06	0.01	0.09	0.95	0.55	0.11	0.03	4.1E-05	49.3
20p11.23	18904507	rs140188658	*C20orf78* [c]	A/G	0.04	-0.02	0.04	0.51	0.72	0.05	-1.63	0.33	2.1E-08	0.53	-0.04	0.02	0.26	95.7
21q21.2	24982945	rs204404	intergenic	A/G	0.15	0.01	0.02	0.76	0.97	0.16	-0.64	0.13	2.1E-07	0.81	-0.005	0.02	0.80	95.9

This table lists the sentinel SNPs (dbSNP147 rs-number) per locus for the loci showing the strongest association with nuclear and cortical cataracts in each study or which contain plausible candidate genes (in bold). Position — chromosomal band and base pair position (GRCh37/hg19, positive strand). Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites). Variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; EA – effect allele; OA – other allele; EAF – effect allele frequency; the effect (beta), its standard error (SE) and p-value are reported; INFO — information score (IMPUTE2 info. for TwinsUK; Minimac R^2 for RSI-III), G – genotyped SNP (INFO=1.00); * SNP was either absent of badly inputed. When a poxy (r^2 >=0.8) that passed QC was available, the estimates for the poxy are presented. The meta-analysis estimates were obtained using inverse-variance fixed effect meta-analysis using METAL. HetI^2 – heterogeneity I^2

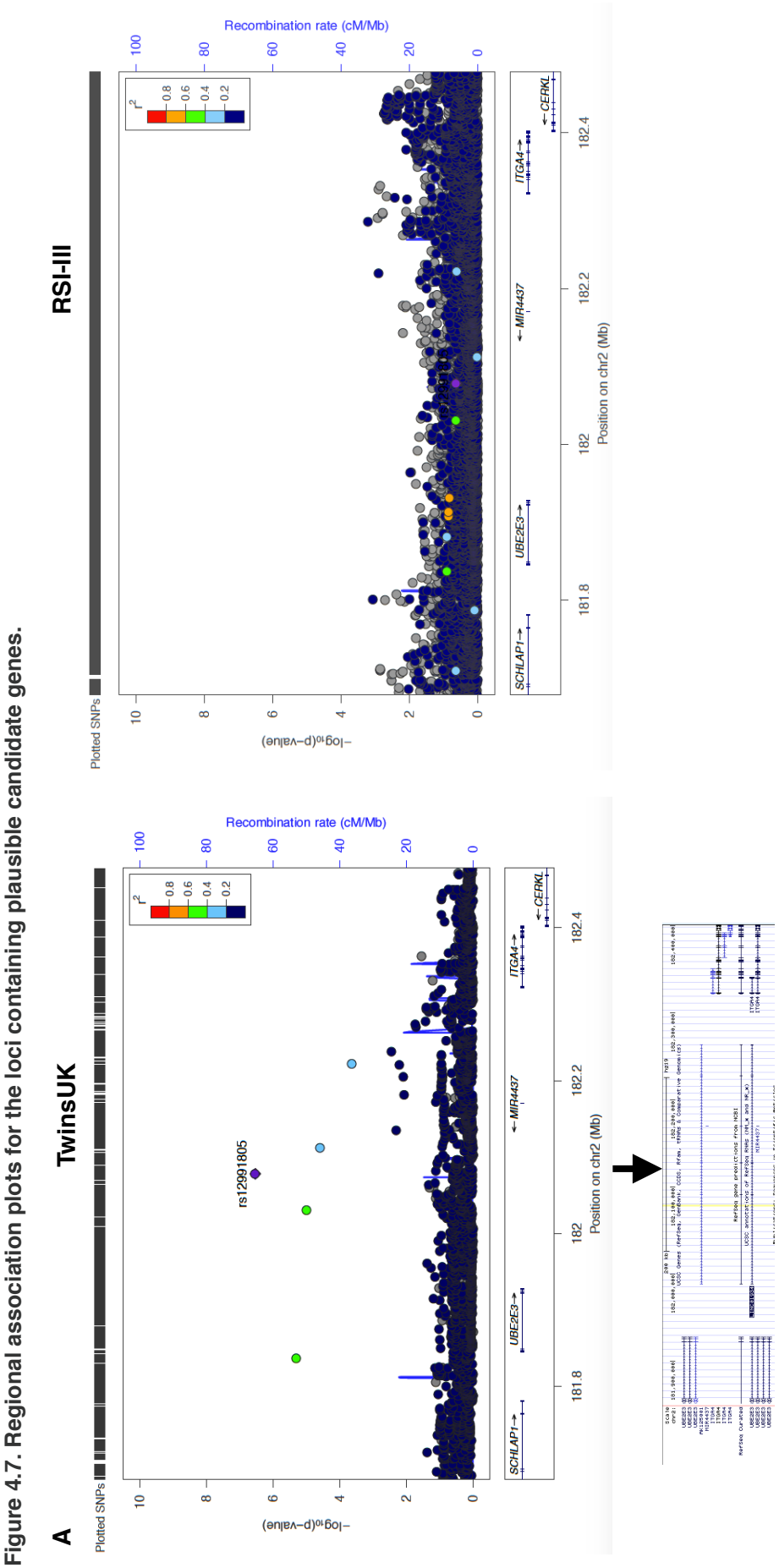
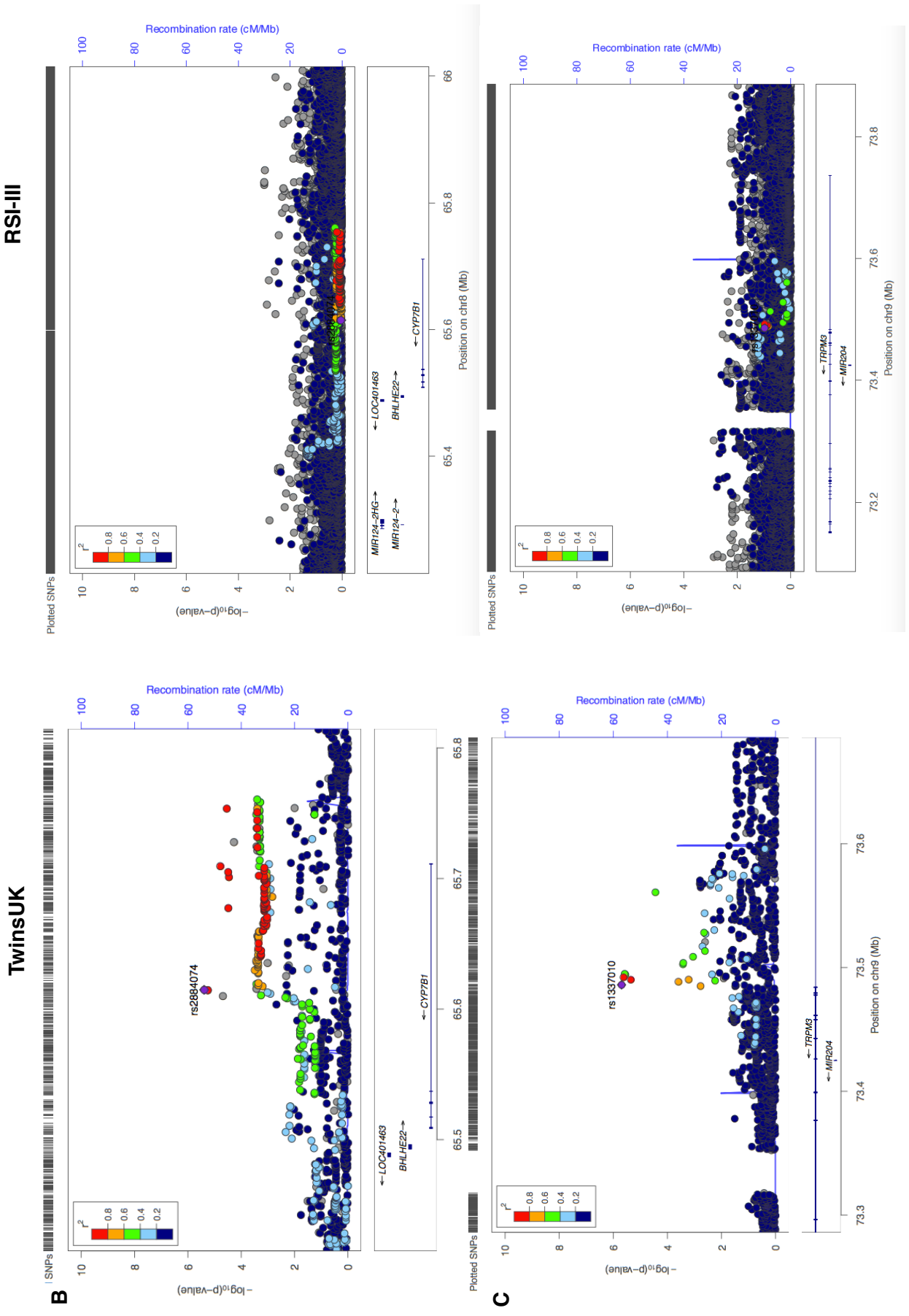
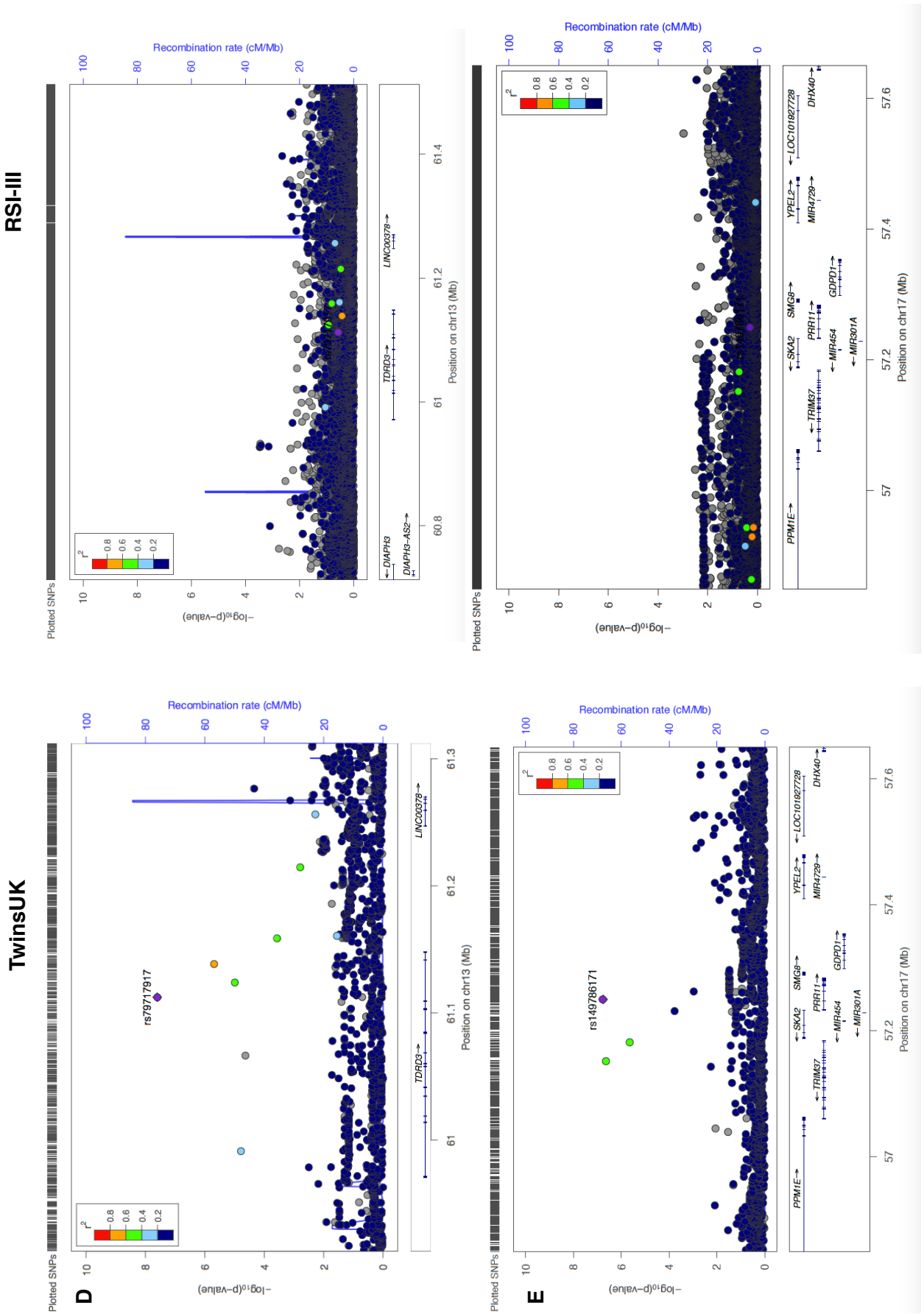
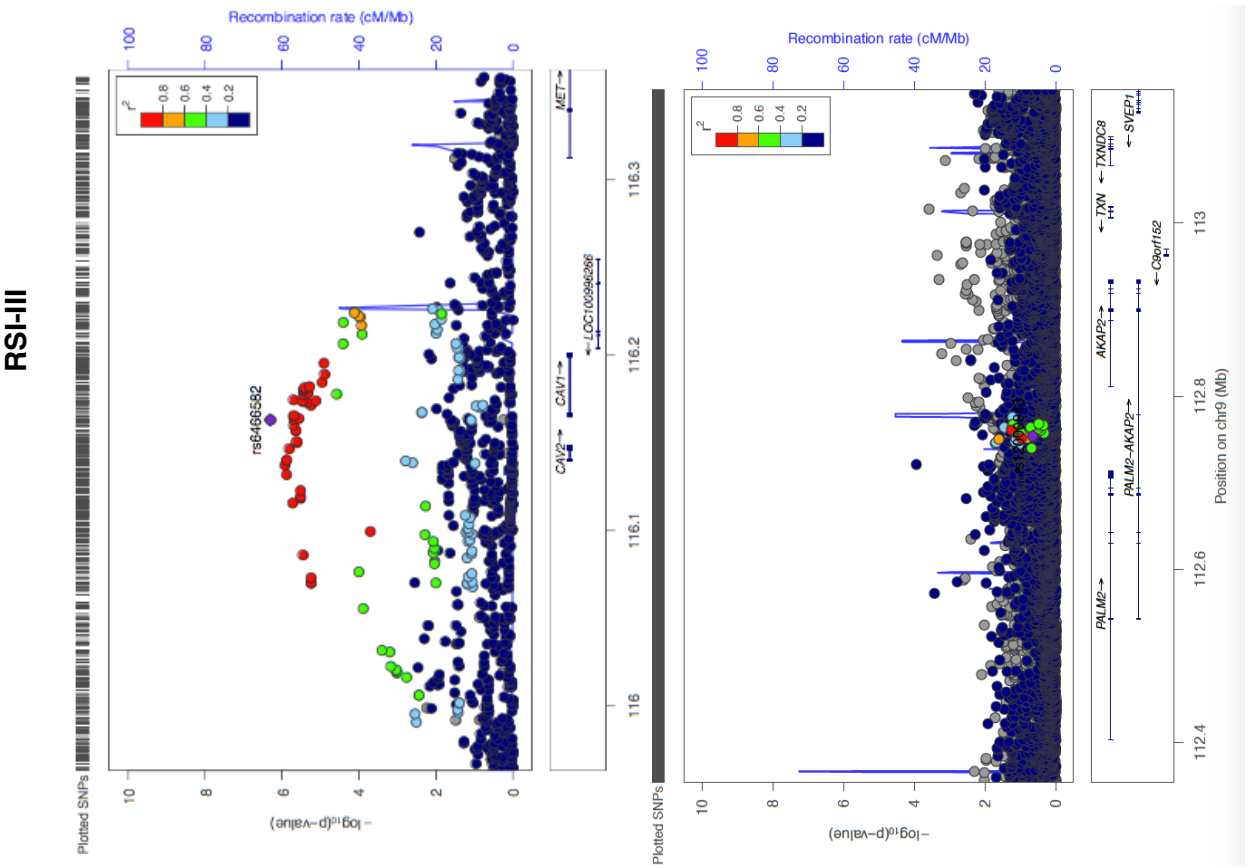
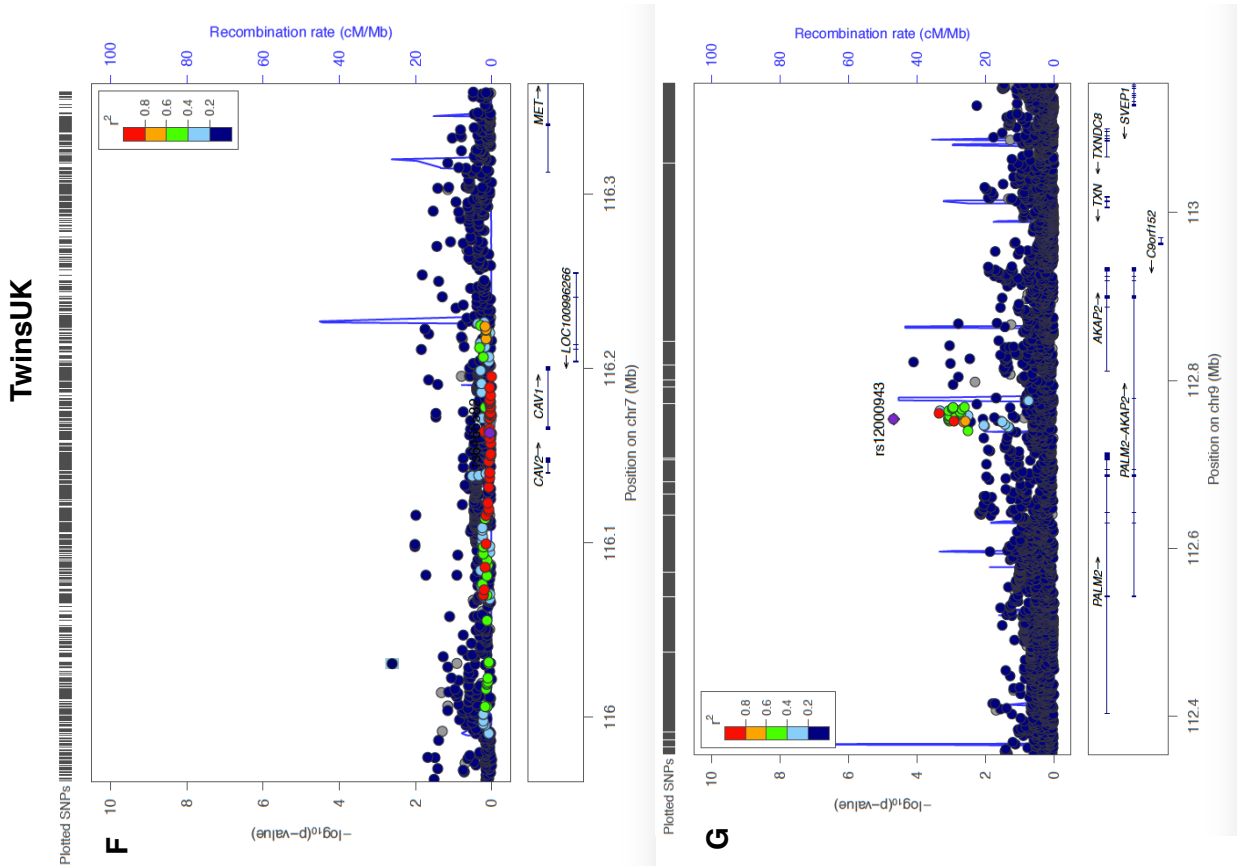


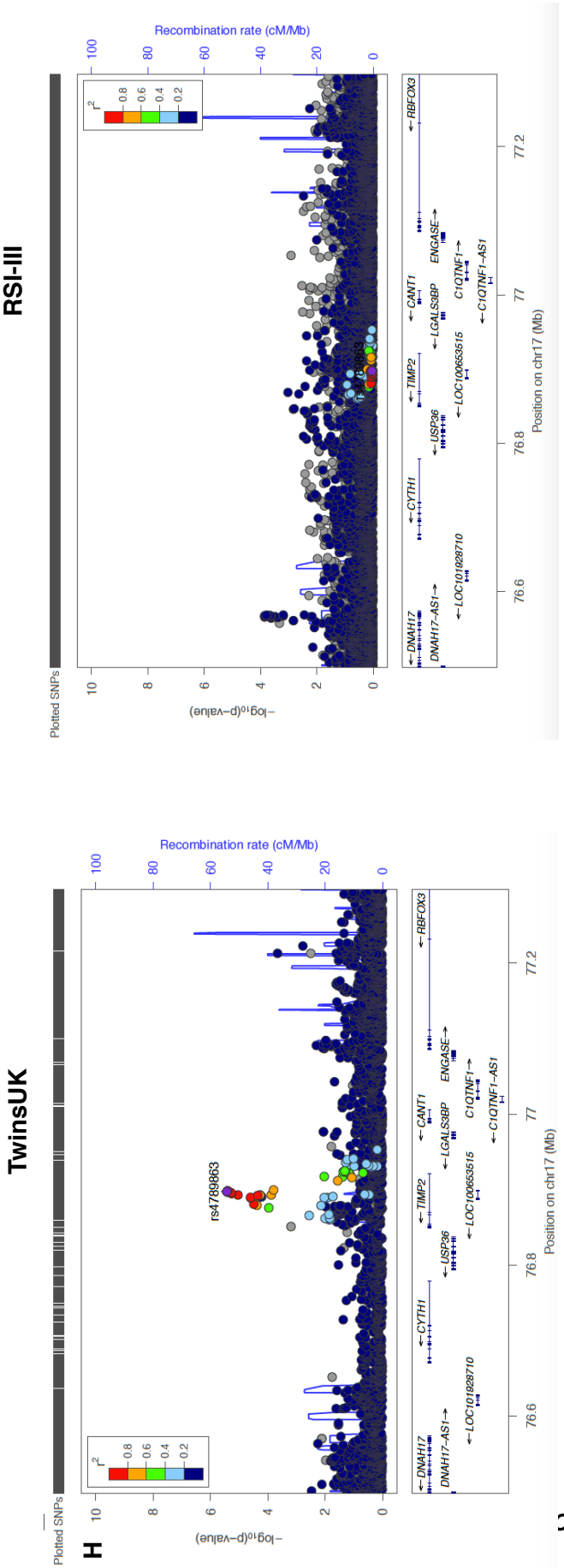
Figure 4.7. Regional association plots for the loci containing plausible candidate genes.


RSI-III
TwinsUK
B
C
Recombination rate (cM/Mb)
−log10(p−value)
Plotted SNPs
SNPs
Position on chr8 (Mb)
Position on chr9 (Mb)
rs2884074
rs1337010
LOC401463
BHLHE22
CYP7B1
MIR124−2HG
MIR124−2
TRPM3
MIR204
r2
0.8
0.6
0.4
0.2

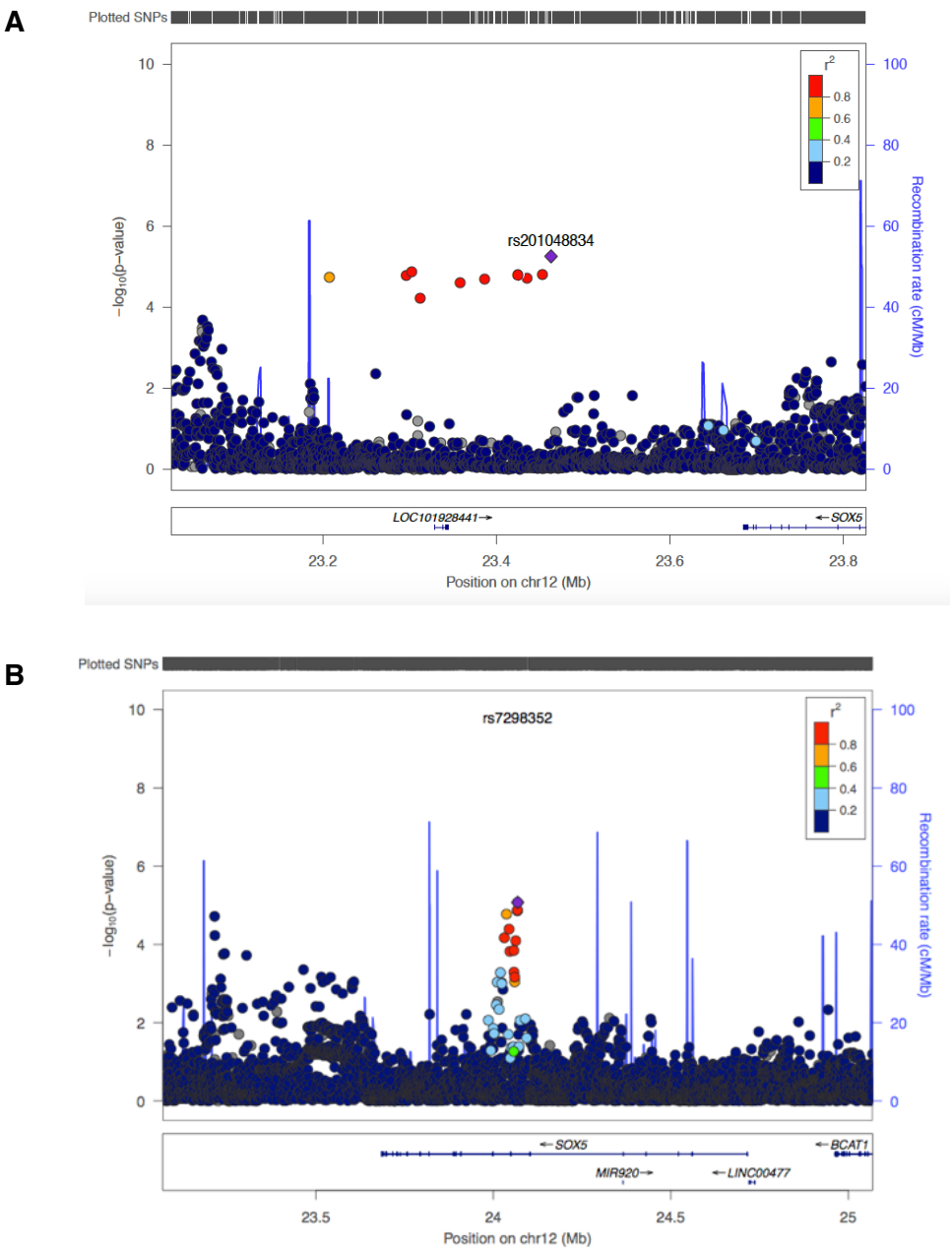

RSI-III
TwinsUK
Recombination rate (cM/Mb)
Plotted SNPs
−log10(p−value)
Position on chr13 (Mb)
Position on chr17 (Mb)
rs79717917
rs149786171
DIAPH3
DIAPH3-AS2
TDRD3
LINC00378
PPM1E
SKA2
TRIM37
SMG8
PRR11
MIR454
MIR301A
GDPD1
YPEL2
MIR4729
LOC101927728
DHX40
D
E


RSI-III
Recombination rate (cM/Mb)
rs6466582
CAV2
CAV1
LOC100996266
MET
Plotted SNPs
−log10(p−value)
Position on chr7 (Mb)
rs12000943
PALM2
PALM2-AKAP2
AKAP2
TXN
TXNDC8
SVEP1
C9orf152
Position on chr9 (Mb)


TwinsUK
F
G
Recombination rate (cM/Mb)
CAV2
CAV1
LOC100996266
MET
Plotted SNPs
−log10(p−value)
Position on chr7 (Mb)
rs12000943
PALM2
PALM2-AKAP2
AKAP2
TXN
TXNDC8
SVEP1
C9orf152
Position on chr9 (Mb)



This figure presents the regional association plots for loci associated with nuclear and cortical cataract in TwinsUK and RSI-III. The plots are centred on the sentinel SNP (purple diamond) and flanked by the GWAS results for SNPs in the 200kb region surrounding it. The shading of the flanking SNPs reflects their pairwise correlation (r^2) with the sentinel SNP. The blue lines represent the estimated recombination rates. A) *LINC01934* locus; B) *CYP7B1* locus; C) *TRPM3* locus; D) *TDRD3* locus; E) *PRR11* locus ; F) *CAV1* locus; G) *PALM2* locus; H) *TIMP2* locus; For the *LINC01934* locus a USCS genome browser snapshot is also added as locus zoom does not correctly display the genes in the region.

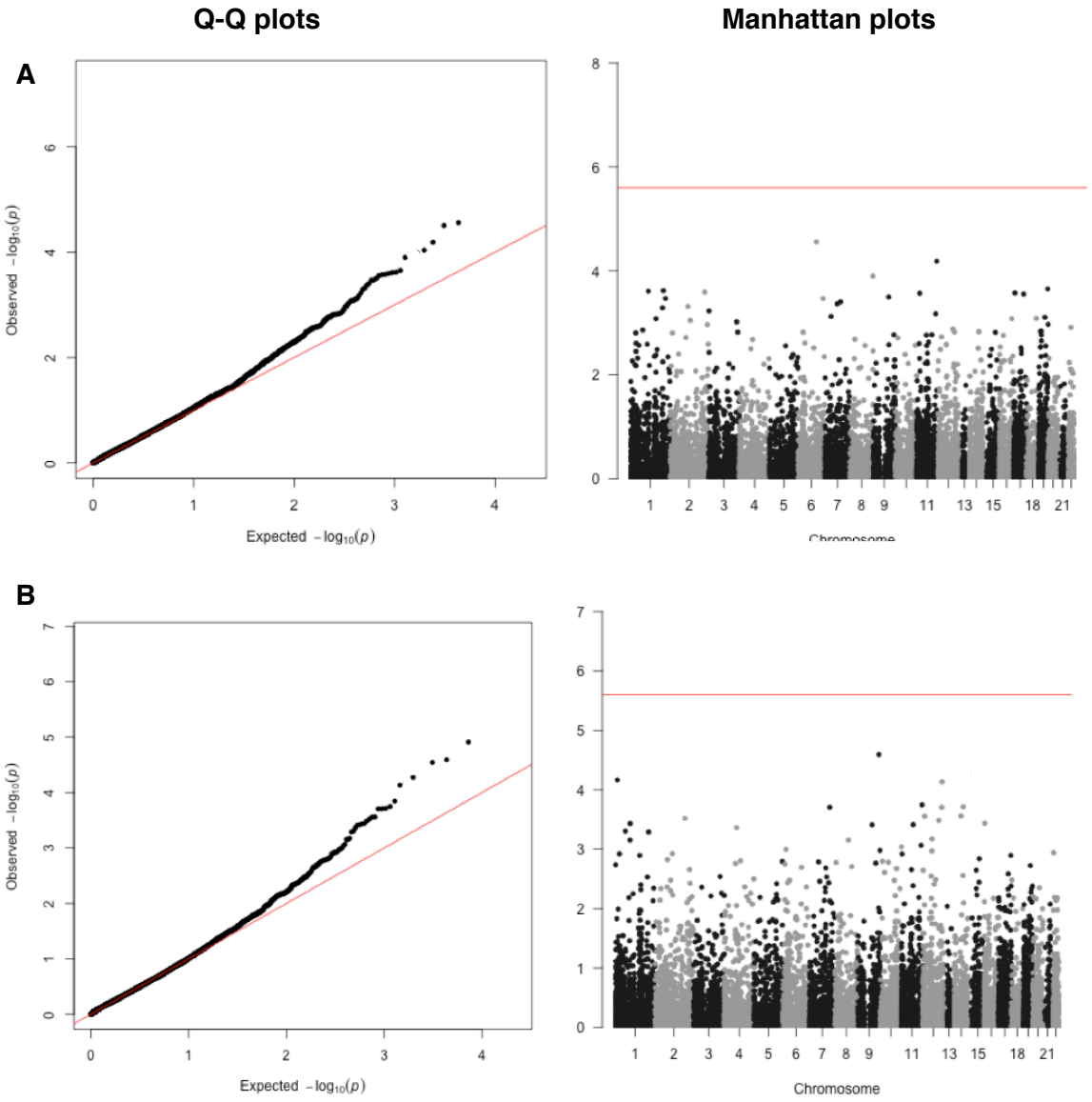


The figure presents the regional association plots for the *SOX5* locus where in TwinsUK variants were associated with both nuclear (A) and cortical (B) cataract. The plots are centred on the sentinel SNP (purple diamond) per locus and flanked by the GWAS results for SNPs in the 500kb region surrounding it. The shading of the flanking SNPs reflects their pairwise correlation (r^2) with the sentinel SNP. The blue lines represent the estimated recombination rates.

4.4.3. Rare-variant analysis in TwinsUK

Here, I present the association results, obtained using SKAT, for the low frequency (MAF<5%) and rare (MAF<1%) exonic genetic variants and nuclear and cortical cataract in 905 unrelated individuals from the TwinsUK cohort. The decision to focus on exonic variants was driven by the relatively small sample available at the time.

Figure 4.9. Quantile-quantile (Q-Q) and Manhattan plots from the SKAT aggregate tests for nuclear and cortical cataract



This figure presents the results from the SKAT aggregate test for nuclear (A) and cortical cataracts (B) in the TwinsUK cohort. Q-Q plots: the observed p-values distribution is plotted agains the expected p-values distribution under the null hypothesis. Manhattan plots: p-values are plotted against the 22 autosomes. red line – genome-wide significance threshold (p-

value=2.5x10^{-6}); blue line – suggestive significance threshold (p-value=1x10^{-5}). All p-values are -log10 transformed. Each dot is a gene.

There was no correlation between the size of the genes, the number of variants analysed per gene or the fraction of individuals harbouring low frequency and rare variants and the p-value obtained from the burden test. There was a low positive correlation (r^2=0.2) between the fraction of people carrying LFRGV in a gene and the gene size, moderate correlation between (r^2=0.4) the number of variants in a gene and the gene size and high positive correlation (r^2=0.7) between number of variants per gene and the fraction of people carrying low frequency and rare variants.

Only exonic non-synonymous variants were considered for this analysis. After QC, 11,263 genes remained and the inflation factors for these analyses were 1.13 in the case of nuclear cataract and 1.12 in the case of cortical cataract (**Figure 4.9, Table 4.5.**). None of the genes reached genome-wide significance and none of the genes associated (P<0.05) with nuclear or cortical cataract has previously been implicated in cataractogenesis.

Table 4.5:SKAT results for low-frequency and rare exonic non-synonymous variants

CHR	BEGINIG	END	GENE	N	FRV (%)	P-value
			Nuclear cataract			
6	116441813	116446506	*COL10A1*	4	2.8	2.8E-05
19	39734352	39735517	*IL28B*	4	5.1	3.1E-05
11	125831724	125871721	*CDON*	6	5.5	6.5E-05
17	56572427	56586149	*MTMR4*	4	1.8	9.2E-05
8	144732092	144732488	*ZNF623*	3	3.4	1.3E-04
			Cortical cataract			
12	57423517	57441459	*MYO1A*	9	10.6	1.2E-05
9	132904231	132904267	*FLJ46836*	3	3.4	2.6E-05
1	228596116	228602816	*TRIM17*	3	1.4	2.9E-05
1	12004663	12032953	*PLOD1*	6	4.3	6.8E-05
11	130716857	130731700	*LOC100507431*	5	4.5	1.8E-04

This table lists the most strongly associated genes when only exonic variants were analysed. CHR — chromosome; beginning and end of longest ORF, gene — gene name; N — number of variants included in final model; FRV% — percentage of people harbouring variants with <5%.

4.5. Discussion

4.5.1. Heritability analysis

In this chapter, I first explored whether nuclear and cortical cataract share genetic risk factors with each other or with other age-related diseases. The classical heritability model yielded results similar to those obtained by Hammond *et al* in a smaller sample of the same population[167,169]. However, the two traits shared only 13% of phenotypic variance and this percentage decreased to 3% when the effect of age was taken into account. There was no statistically significant sharing between the two traits irrespective of the model used. A bigger sample size is, therefore, needed to accurately estimate any shared heritability, given such a low phenotypic sharing.

The classical heritability modelling uses phenotypic data only and, apart from partitioning the variance attributed to genetic factors into variance due to additive and variance due to dominant effects, it does not provide in-depth information about the genetic architecture of a trait (number of loci involved, strength of effects at a locus). SNP-based heritability methods are better suited for elucidating the genetic architecture. In line with the "missing" heritability problem[495], both SNP-based tests yielded lower heritability estimates than the classical twin model. However, GCTA and LD-score regression gave very different results even though the same variants were used for both types of analysis. Namely, using LD score regression resulted in much lower heritability. This results is in line with previous studies where the LD-score regression estimates are consistently smaller than those from GREML in the same data set[523]. In addition, using different sets of SNPs (data not shown), I obtained vastly different results from no shared heritability to very high estimates which means that the GCTA results are very dependent on the sets of SNPs chosen. Indeed, it has been previously shown that when a sparser set of variants is used as opposed to a denser one, a GREML analysis resulted in a smaller proportion of phenotypic variance explained[524].

Using GREML may lead to biased estimates when significant part of the effect is explained by rare causal variants, which are unlikely to be well tagged by common SNPs, or when causal variants are located in genomic regions with a different LD properties from the rest of the genome[524]. LD-score regression is more robust to varying LD structure but it is less well suited to deal with rare variants. Furthermore, GCTA-GREML calculates LD from within the sample while LD-score regression uses a reference sample to do that. Another limitation of LD-score regression is that its results

may be less stable when estimated for smaller samples (overall sample size<50,000). Here, it is important to point out that for both the GCTA and LD-score regression analyses, only a subset of SNP were used (HapMap3 SNPs) and, given that the TwinsUK was imputed to the 1000G panel, this subset may not capture the full extent of the genetic effects from the additional non-HapMap3 SNPs. In both cases, using more variants and a better reference population, such as UK Biobank, to calculate LD may improve the estimates. In addition, models such as LDAK* or GCTA-LDMS may afford a better estimation of the underlying LD-architecture[525,526].

In conclusion, probably the true shared heritability between nuclear and cortical cataract lays somewhere between the GCTA and LD-score regression. GCTA is probably overestimating[527] while the LD score regression is probably underestimating the heritability, but in any case the shared heritability between the two traits seems to be still low. This conclusion is still supported by the fact that the polygenic risk score analysis showed exceedingly small sharing of genetic risk factors. Thus, nuclear and cortical cataract seem to be genetically different diseases, which is interesting given they often occur together, but this is similar to the environmental associations which, other than age, are also different.

4.5.2. Genome-wide association analysis

In this chapter, the first GWAS results for nuclear and cortical cataracts in two European ancestry cohorts have also been presented. The majority of sentinel variants were of low frequency and showed a substantial heterogeneity of effects between the two cohort. Thus, the two cohorts largely did not replicate each other. Given that both cohorts were underpowered these results are not surprising. Meta-analyses of the results presented here together with GWAS results from other cohorts would afford the better power needed to determine any true association which much greater statistical confidante. Despite that, 2 of the variants associated with nuclear cataract show evidence for association in both of the cohorts. Furthermore, some of genes within the associated loci (p-value < $1x10^{-5}$) may be interesting functional candidates for nuclear or cortical cataract formation and will be discussed in detail here. As explained in the introduction of this chapter (**4.2.2. Genome-wide association studies**), no fine-mapping was attempted here, but candidate genes were shortlisted based on the following information present in the publicly available literature: the function of the

* LD-adjusted kinship

protein they encode when known; expression patterns concordant with lens or eye expression; interaction with genes previously implicated in cataract formation; the gene being member of a gene family where other members have been previously implicated in cataract formation.

***LINC01934* and *PRR11* loci:**

The association between the sentinel variants at both loci and nuclear cataract was supported by both cohorts (rs12991805 within *LINC01934*, meta-analysis p-value $7.5x10^{-9}$; rs149786171 within *PRR11*, meta-analysis p-value $1.2x10^{-9}$). In both cases, the sentinel variant mapped to a region of extended LD (**Figure 4.7.**) but, given their low frequency (7% and 1% respectively), the variants themselves were not in strong LD with other variants in the region. On bases of their function and expression, none of the genes at the two loci were strong candidate genes. Given that lncRNA my be regulating gene in trans and that there are a couple of microRNA genes in proximity to the *PRR11* sentinel variant, it may be that the association, if real, is due to target genes in trans.

***CYP7B1* locus:**

The sentinel variant within the *CYP7B1* locus (rs2884074, $p\text{-value}_{TwinsUK} = 4.1x10^{-6}$) mapped to the 5'UTR of one of *CYP7B1* isophorms. Similarly to the *LINC01934* locus, the *CYP7B1* sentinel variant was situated in a long LD-block (**Figure 4.7.**). Thus, although the pick was located over *CYP7B1* it is unclear whether *CYP7B1* is the causative genes. Strong LD with a relatively common allele in a relatively large region may indicate a recent mutation event under positive selection[528]. Furthermore, in terms of function, *CYP7B1* is a likely candidate. The genes takes part in cholesterol metabolism, which may be important in cataract development[529], and Mendelian mutations in *CYP7B1* cause spastic paraplegia in combination with cataracts[530].

***TRPM3* locus:**

In the case of the *TRPM3* locus(rs1337010, $p\text{-value}_{TwinsUK} = 2.0x10^{-6}$), the *TRPM3* gene and its microRNA were the only two genes in LD with the sentinel variant (**Figure 4.7**). *TRPM3* and its regulating microRNA (*MIR204*) are expressed in lens and lens epithelium; and are under the control of PAX6 — a major eye-development transcription factor[531-533]. TRPM3 is a membrane-bound calcium channel, which was previously linked to congenital cataracts[534-536]. The *Mir204* is a negative regulator of gene expression (including of *Sox11*) in vertebrate eye[537].

***TDRD3* locus:**

Similarly to the *TRPM3* locus, the *TDRD3* gene was the only gene in LD with the sentinel variant (rs79717917, $p\text{-value}_{TwinsUK} = 2.5x10^{-8}$) at this locus (**Figure 4.7**). *TDRD3* gene encodes for a scaffolding protein in cytoplasmic stress granules[538].

Mutations in a gene (*TDRD7*) encoding for another member of the same family of proteins, caused paediatric cataracts[539]. Loss-of-function mutations in the murine *Tdrd7* lead to cataracts and glaucoma[539].

***CAV1/2* locus:**

A sentinel variant (rs6466582, p-value$_{RSI-III}$ = 5.0x10^{-7}) in the *CAV1/2* locus showed suggestive association with cortical cataract. This locus showed similar LD-structure to the *CYP7B1* locus (**Figure 4.7**). One of the genes in the locus, *CAV1*, is well studied in relation with age-related eye diseases. For example, common *CAV1* variants have been associated with primary open angle glaucoma[540]. Although the role of *CAV1* in cataract formation is unclear there is some evidence suggesting that the gene may be involved. *CAV1* encodes for a scaffolding protein expressed in most cells types, including lens fibres. Moreover, knock-out *Cav1* mice develop cataract[541] and heterozygous de-novo *CAV1* nonsense mutations have been proposed to cause neonatal lipodystrophy with cataract[542].

***PALM2/ AKAP2* locus:**

In the *PALM2/AKAP2* locus there seem to be two independent signals coming from two separate LD-block (**Figure 4.7**). This locus has a complex structure and there are naturally occurring read-through products of the *PALM2* and its neighbouring gene *AKAP2*. Both genes have been implicated in lens function. In mice, *Palm2*, a gene proposed to be under *Pax6* control, is expressed in the lens placode, and this expression is maintained in the lens vesicle throughout the formation of the adult lens[543]. In vertebrate, AKAP2 couples MIP to its protein kinase supporting support ocular lens transparency[544].

***TIMP2* locus:**

At the *TIPM2* locus there are more than 10 genes located in the same LD-block (**Figure 4.7**) where the sentinel variant maps to (rs4789863, p-value$_{TwinsUK}$ = 3.7 x10^{-6}). However only one of the 10 genes (*TIMP2*) has a known eye-related function. *TIMP2* encodes for an aqueous humor metalloproteinase inhibitor that has been proposed to have a relationship with in glaucoma and axial length[545,546]. Matrix metalloproteinases are known mediators of various cataract types[547].

Even though the above discussed genes are plausible candidates, the two studies largely did not replicate each other. A meta-analysis of those and other available studies would afford better power to detect genes that are truly associated with ARC. Such an analysis is presented in the following chapter.

Finally, neither the TwinsUK or the RSI-III GWAS analyses detected an association between variants in *EPHA2* the only gene, to date associated with age-

related cortical cataract[26,122,208,209]. There is a simple explanation for that result, namely the very different method used between the present and the past *EPHA2* studies. Here I have used all individuals that had a cataract score and a case-control cut-off at 5%. The method I have used here is better powered to detect common variants associated with cataract and assumes that cortical cataract is a common polygenic trait.

The association between the *EPHA2* locus and cortical cataract was initially discovered using linkage analysis and candidate gene/variants studies[26,122,208,209]. The *EPHA2* variants associated with ARC were found using cut-offs for defining case and control status 25% of lens area affected and <1% of area affected respectively[122]; effectively cutting off the middle of the score distribution and giving more weight to the strongly affected individuals. Such a method is also a way to enrich the right tail of the distribution for Mendelian forms of cataract. Indeed, when 25% cut of is used, EPHA2 is associated with cortical cataract in both TwisnUK and the RSI-III (rs7548209, p-value of $2x10^{-4}$ and $1x10^{-3}$ respectively). Using a similar study design may reveal variants in additional genes with moderate to strong effects but will also necessitate much bigger sample size.

4.5.3. Low-frequency and rare variants analysis

Finally, I also explored the effect LFRGV have on ARC in TwinsUK. As pointed out in the methods section of this chapter, rare-variants are more susceptible to the effects of even subtler population structure[517-519]. Thus, even though the WGS was done in a set of unrelated individuals, I opted for using a kinship matrix. This indeed resulted in substantially reduced inflation but the results were still inflated and if genomic control was used, none of the genes would be significant at suggestive level (p-value < $1x10^{-5}$). Moreover, none of the genes previously linked to congenital cataract were associated with ARCs (p-value > 0.05). I decided to focus on exonic variants, as the sample size of 905 individuals is not sufficient to yield power for whole-genome analysis. Making use of the additional WGS samples from the HLI TwinsUK collaboration (**Table 3.1**.) when they become available would increase the sample size and will make it possible to explore other groups of functional variants (eg. within regulatory regions or non-coding RNA classes). Exploring better powered and more versatile tests such as adaptive mixed effect models implemented in MONSTER[548] is also worthwhile. In addition these and any other results obtained by other methods will need to be replicated. The RSI-III has WGS data available which, in combination with the fact that phenotyping in RSI-III was done in the same way as TwinsUK, makes this

cohort a good replication set. Even then, however, replication may prove problematic as rare variants may be population specific and thus difficult to replicate in other populations.

4.5.4. Conclusions

In conclusion, the shared heritability analysis in this chapter showed that nuclear and cortical cataract seem to be genetically different conditions. Those results also were supported by the fact that the loci that were found to be associated with nuclear and cortical cataract in two European ancestry cohort did not overlap. Better powered studies will be needed in order to confidently detect cataract-associated loci and to fine-map causative variants and genes. Finally, low-frequency and rare genetic variants within coding regions of the genome did not cumulatively show statistically significant effects on cataract formation. Again, a better power study will afford better power to explore the cumulative effect of non-coding variants, which are much more likely to affect a common complex phenotype such as cataract.

Chapter 5. Meta-analysis of genome-wide association studies of age-related nuclear cataract

5.1. Overview

The power of GWAS is heavily dependent on the sample size and, as seen in our own data in the previous chapter, small studies are underpowered to find associations which would survive correction for multiple testing. In order to increase power and reduce false-positive findings, meta-analyses are performed. As described in **Chapter 1**, a meta-analysis of GWAS for nuclear cataract in Asian populations led to the identification of the first two loci for age-related nuclear cataract[27]. Here, I perform the first meta-analysis of GWAS for nuclear cataract in cohorts of European ancestry and combine those with GWAS of Asian ancestry cohorts. First, however, I will briefly describe the concept of meta-analysis in genetic studies and the methods traditionally used to perform those.

5.2. Introduction

Meta-analyses are the statistical synthesis of summary-level information from multiple independent studies, and have dramatically increased the yield of discovered and validated genetic risk loci[549]. Commonly, meta-analyses are performed using effect estimates weighted by the inverse of the standard error so that more weight is given to studies with more precision[550]. More imprecise, sample-size-weighted p-value meta-analyses are also possible[550] and can be used to combine results from a mix of quantitative and binary analysis. Meta-analysis usually uses frequentist (fixed or random effect) methods. The fixed effect meta-analyses assume a single common effect underlying all the studies, while random effects analysis makes the assumption that individual studies are estimating different underlying effects[550]. When between-study heterogeneity is absent, fixed and random effects calculations yield identical point estimates and confidence intervals. With increasing between-study heterogeneity, the random effects summary estimates have larger variance (wider confidence intervals) and usually, but not always, they have lower statistical significance[550]. This drawback is due to overly conservative null-hypothesis model that assumes heterogeneity instead of no association (effect size of 0) under the null hypothesis, and can be amended by using Han and Eskin's random effects model[551]. The test replaces the traditional z-score estimation of p-values with a likelihood-ratio test and assume no heterogeneity under the null hypothesis[551].

In practice, a mixture of fixed effects rather than a true random effect is more likely[552]. In this case, Bayesian models, such as the ones applied in trans-ethnic

analysis, may be more suitable. These models apply specific prior probabilities to the uncertainty parameters and one can examine the robustness of conclusions based on these priors, such as whether different priors may change the results perceptibly[550]. In this chapter both frequentist and Bayesian methods will be explored.

Similarly to meta-analysis in epidemiology, heterogeneity of effects in genetic meta-analysis presents a challenge as it decreases the power to detect true effects and increases the likelihood of detecting false positives[553]. Moreover, in presence of heterogeneity the effects are smaller in size and that may partially explain the discrepancy between the classically calculated heritability and the proportion of variance explained by associated genetic variants[553]. Heterogeneity, may result from methodological errors such as: population stratification, different study designs with differential ascertainment of phenotype across cohorts, the phenotype of interest being correlated with another phenotype with which the SNP is associated, differential genotyping errors, or imputation errors leading to breaking of LD patterns between tagged markers and causal variants across cohorts[554,555]. In these cases, heterogeneity may lead to false positive findings and it is extremely important that it is measured and dealt with. Between-cohort heterogeneity in GWAS effects may reflect genuine differences between cohorts that are results from local gene-by-environment interactions or, in multi-ethnic studies, from differences in LD[556,557]. Exploring heterogeneity in this case can bring insight about the underlying biology. There are different methods to test for and measure the amount of between-study heterogeneity. The two most commonly used are Cochran's Q and I^2 statistics. Cochran's Q measures the weighted differences between the summary effect size and the study-specific effect sizes, and it is typically considered significant at the $\alpha = 0.10$ level[554]. This test is underpowered, when there are very few studies or when the studies are small studies with large confidence intervals[554]. I^2 is a measure of the total variation across studies due to heterogeneity beyond chance and is calculated as: (Q-(number of studies - 1))/ Q). This metric is independent of the number of studies and can be compared across meta-analyses with different numbers of studies and metrics[558,559]. The second method is the one I have used when performing the meta-analysis.

As with GWAS, meta-analysis of GWAS can identify associated loci but not the causative variants or genes. Therefore, in order for causal variants or candidate genes to be identified and prioritised for functional in vitro experiments, meta-analysis is followed by bio-informatics analysis of the identified loci. First, the association picks can be refined using conditional analysis (GCTA) or credible set analysis (FINEMAP, CAVIAR), meaning one variant or a small subset of variants are identified that are most

likely to be the causative variants or to tag a unobserved causative variant[488,560,561]. If data from samples with different ethnic backgrounds is available, trans-ethnic meta-analysis can also help to identify potentially causative variants. This type of analysis relies on the fact that if an associated allele is shared between groups with different ancestry, can be useful to help fine-map a locus by restricting the credible set of variants to those that are in LD with the sentinel variant in all populations[562].

The potential functional consequences of the identified variants or a set of variants can next be explored. If the variants map to coding regions their effect on protein function or its stability can be predicted using tool such as polyphen2, SIFT and CADD which predict likely structural and functional consequence and in addition incorporate data on evolutionary site conservation[563]. Variants are more likely to map within non-coding regions in which case one needs to explore whether they are disrupting regulatory sequences for or disrupt expression in target tissues (expression quantitative trait loci (eQTL) analysis) of any of the neighbouring genes (cis-effects)[563]. Additional levels of evidence such as gene-set enrichment analysis, expression patterns of genes in target tissues in humans and model organisms, co-expression and co-binding analysis can also be very helpful when prioritising candidate genes[563]. To better understand the functional connection between the meta-analysis identified loci and age-related nuclear cataract, some of the here described methods are employed in this chapter.

5.3. Materials and Methods

5.3.1. Cohorts of the International Cataract Genetics Consortium

Phenotyping, genotyping and imputation procedures used in the different cohorts of the International Cataract Genetics Consortium (ICGC) have been discussed in detail in **Chapter 2** and summarised in the two tables below (**Table 5.1.** and **Table 5.2.**).

5.3.2. Study design

Age-related nuclear cataract GWAS summary statistics from 12 studies comprising 19,776 individuals were meta-analysed (**Table 5.1.**). This was done in 3 phases: discovery (N=14,151), replication (N=5,625), and all data meta-analysis (**Figure 5.1.**). As LD patterns and causative variants between ancestry groups may

differ, in the discovery phase individuals from European (N=7,352) and Asian (N=6,799) ancestry were analysed separately, followed by combined analysis.

Table 5.1: Clinical characteristics and cataract phenotyping of participating cohorts

Cohort	Ancestry	N	Age, yrs (Mean ± SD)	Sex (% F)	Phenotyping and grading methods	NC grade (mean ± SD)
Discovery (n = 14,151)						
AREDS	European	1434	67.4 ± 4.7	59.0	LPh, AREDS	2.2 ± 0.8
BMES	European	2209	66.3 ± 8.8	57.1	LPh, Modified Wisconsin	2.5 ± 0.7
RSI-III	European	1424	71.2 ± 6.2	54.1	LPh, CNDS	89.6 ± 36.9
TwinsUK	European	2285	62.2 ± 6.8	98.0	LPh, CNDS	62.1 ± 18.1
SCES-610	Asian (Chinese)	1704	57.5 ± 8.9	48.8	LPh, Modified Wisconsin	2.3 ± 0.7
SCES-OmniExpress	Asian (Chinese)	526	58.7 ± 8.5	48.3	LPh, Modified Wisconsin	2.4 ± 0.7
SiMES	Asian (Malay)	2369	58.4 ± 10.9	50.6	LPh, Modified Wisconsin	2.5 ± 0.7
SINDI	Asian (Indian)	2200	56.4 ± 9.1	48.3	LPh, Modified Wisconsin	2.3 ± 0.7
Replication (n = 5,625)		**N (cases)**				
BDES	European	1509 (280)	62.3 ± 10.5	53.7	LPh, Original Wisconsin	2.7 ± 0.9
BES	Asian (Chinese)	1528 (769)	57.4 ± 9.4	61.6	LPh, AREDS	2.6 ± 1.1
INDEYE(S)	Asian (Indian)	2284 (373)	66.7 ± 6.0	52.4	LPh, LOCS III	3.7 ± 1.1
SLCCS	European	304 (242)	73.9 ± 10.2	54.9	SLB, LOCS III	2.4 ± 0.7

This table summarises the clinical characteristics of the participant in each cohort and the phenotyping methods used. AREDS – Age-related Eye Diseases Study (and its grading system), BMES – Blue Mountains Eye Study, RSI-III – Rotterdam Study I (follow-up 3), SCES – Singapore Chinese Eye Study; SiMES – Singapore Malay Eye Study; SINDI – Singapore Indian Eye Study; BDES – Beaver Dam Eye Study; BES – Beijing Eye Study; INDEYE – The India Eye Study; SLCCS – South London Case Control Study; N — number of individuals studied, the number of case for the replication cohorts is given in brackets after the total number of individuals; % F — percentage of females; SD — standard deviation; LPh — lens photography, SLB — slit lamp bio-microscopy; CNDS — central nuclear dip score; Wisconsin— Wisconsin Cataract Grading System; LOCS III — Lens Opacities Classification System III;

In order to maximise power in the discovery phase (N = 14,151), I selected for meta-analysis the GWAS where quantitative nuclear cataract scores were available. In these GWAS, all participants underwent detailed eye examination including lens photography after pupil dilation, which permitted accurate grading of cataract from lens

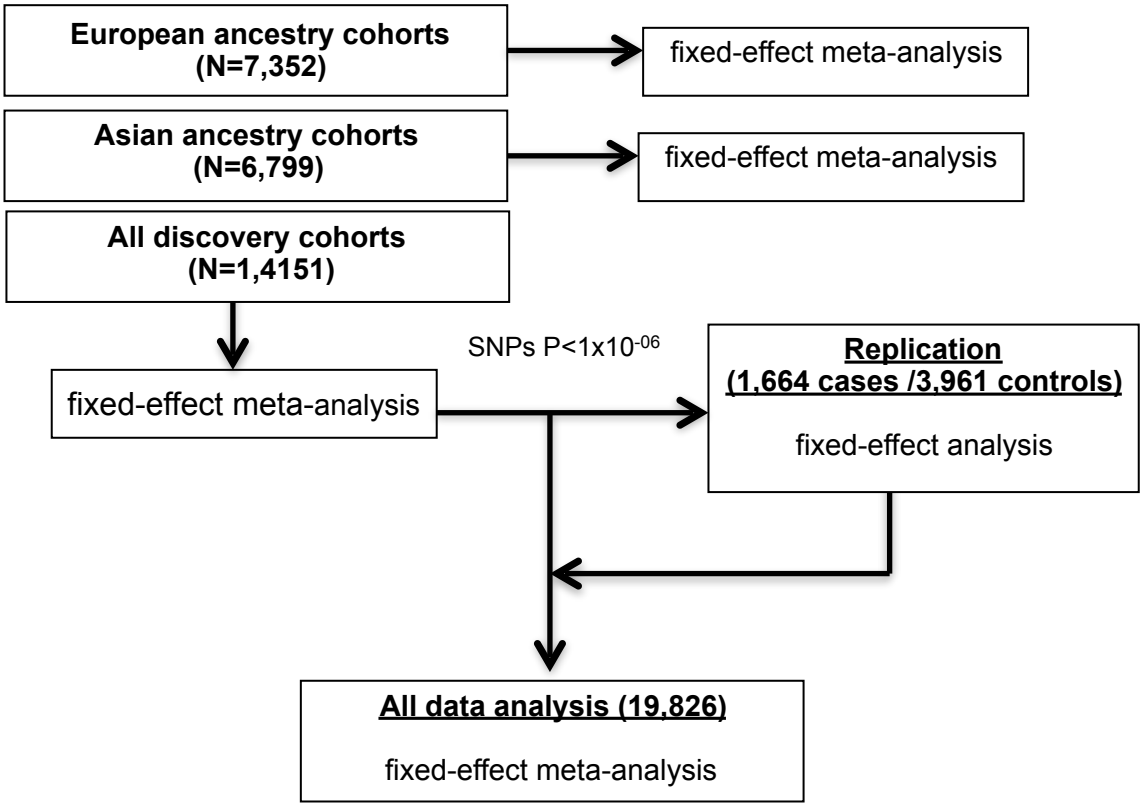
photos using standardised grading that provided decimalised grades. Moreover, all studies used SNPs imputed to the 1000 Genomes panel (**Table 5.2.**). Each study included in the discovery phase performed per-SNP linear regression (additive genetic model) analysis separately following a pre-agreed protocol. All studies adjusted for age, sex and, when necessary for principal components of population structure. In addition, relatedness was taken into account in the TwinsUK cohort. Furthermore, prior to the association analyses nuclear scores were, when necessary, normalised. As different cataract grading systems were used, cataract scores were standardised within each cohort by subtracting the mean and dividing by the standard deviation.

Table 5.2: Genotyping and imputation platforms in the participating cohorts

		1000G Imputation	
Cohort	**Genotyping platform**	**Software**	**Reference Panel**
Discovery			
AREDS	Illumina HumanOmni2.5−4v1_B chip	IMPUTE2	Phase 1, ALL 2012
BMES	Illumina Human 670-Quadv1 custom	IMPUTE	Phase 1, EUR 2012
RSI-III	Illumina Infinium II HumanHap550 v3.0	Minimac	Phase 1, EUR 2012
TwinsUK	Illumina HumanHap 300kDuo & HumanHap610-Quad	IMPUTE2	Phase 3, EUR 2014
SCES-610	Illumina HumanHap610-Quad	Minimac	Phase 1, ALL 2012
SCES-OmniExpress	Illumina OmniExpress	Minimac	Phase 1, ALL 2012
SiMES	Illumina HumanHap610-Quad	Minimac	Phase 1, ALL 2012
SINDI	Illumina HumanHap610-Quad	Minimac	Phase 1, ALL 2012
Replication			
BDES	Illumina Human Exome Array	Minimac	Phase 1, ALL 2012
BES	Illumina HumanHap610-Quad	Minimac	Phase 1, ALL 2012
	Illumina OmniExpress	IMPUTE2	Phase 3, ALL 2014
INDEYE(S)	TaqMan	-	-
SLCCS	Illumina OmniExpress Exome Array	Minimac	Phase 3, ALL 2014

The table summarises the genotyping and imputation procedures used by the ICGC cohorts. AREDS – Age-related Eye Diseases Study (and its grading system), BMES – Blue Mountains Eye Study, RSI-III – Rotterdam Study I (follow-up 3), SCES – Singapore Chinese Eye Study; SiMES – Singapore Malay Eye Study; SINDI – Singapore Indian Eye Study; BDES – Beaver Dam Eye Study; BES – Beijing Eye Study; INDEYE – The India Eye Study; SLCCS – South London Case Control Study; EUR – European ancestry; ALL – all populations in 1000 Genome project.

Figure 5.1. Meta-analysis flow-chart



This graph shows the design of the current meta-analysis, the type of analyses and the number of individuals used at each stage.

The most strongly associated SNPs in each of the discovery phase loci that were associated with age-related nuclear cataract (p-value < $1x10^{-6}$) were selected for replication in four additional independent cohorts (N=5,625; 1,664 cases and 3,961 controls). In this phase, the cohorts where the phenotyping was less accurate (BDES, BES and SLCCS) or which did not have genome-wide data (INDEYE) were used. Because of the less accurate phenotyping, it is problematic to include discrete scores as truly quantitative variables in regression models. Therefore, in the replication phase age-related nuclear cataract was treated as a binary outcome and logistic regression was performed. Case status was determined on the basis of the grading system used in each replication study as follows: BES (AREDS grade ≥ 3), INDEYE(S) (LOCS III grade ≥ 3), BDES (Wisconsin grade ≥ 4), and SLCCS (LOCS III grade ≥ 3). Controls were individuals free of any nuclear opacity and cataract surgery. Finally, in the all data meta-analysis (N=19,776), discovery and replication cohorts were meta-analysed together.

5.3.3. Quality control and statistical analysis

The quality control of the summary statistics from the separate GWAS was done using a protocol developed by the Genetic Investigation of Anthropometric Traits (GIANT) Consortium[516], as described in detail in **Chapter 4** (section 4.3.3. Statistical analysis; subsection: GWAS results QC). Briefly, all variants with MAF <1%, information score of <0.3, and HWE p-value $<1x10^{-4}$ were excluded. As data on insertion-deletion polymorphisms was missing for some of the cohorts, single base changes only were analysed. Variants present in less than half of the cohorts per ancestry group, or in less than 3 of the cohorts in the combined analysis were also excluded.

After QC, between study heterogeneity was calculated using METAL[522]. As the vast majority of variants showed low to moderate heterogeneity, I opted for a fixed effect model for all stages (METAL[522]): inverse-variance (discovery, replication); sample-size-weighted p-value (all data meta-analysis). As substantial genomic inflation can be present even in the absence of population substructure or technical artefacts[564], no genomic correction was applied.

In the discovery phase additional analyses were carried out. For validation purposes, a random effect meta-analysis (METASOFT[565]) and sensitivity analysis were performed — excluding cohorts one at a time and re-running the meta-analysis. To evaluate sharing between European and Asian Cohorts, trans-ethnic meta-analysis (MANTRA[566] and MRMEGA*) were performed. Finally, to refine the association loci, conditional analysis using GCTA[488] was performed as follows. At each locus the least-squares estimates of the effects of SNPs at the locus conditional on the most strongly associated independent SNP at the locus were calculated. The 1000G Phase3 data was used to calculate LD patterns at each locus. In this analysis, I explored the effect of using different ancestry panels (mixed, European, East Asian, South East Asian). GCTA was also used to calculate the amount of genetic variance explained by associated SNP[488]. Shared heritability between age-related nuclear cataract and other traits for which GWAS results were available through http://ldsc.broadinstitute.org/, and heritability partitioning were explored using linkage disequilibrium score regression[489]. Finally, gene-based meta-analyses (GATES[567]) were also performed.

* http://www.geenivaramu.ee/en/tools/mr-mega

5.3.4. Gene-set enrichment analysis

Gene-set enrichment analysis were performed using MAGENTA[568]. For this analysis GWAS meta-analysis variants with p-value less than the following cut-offs: 0.05; $1x10^{-3}$; $1x10^{-4}$; and $1x10^{-5}$, were selected and analysed separately. The number of permutations used to calculate the false discovery rate (FDR) adjusted p-value was 10,000 and, as nuclear cataract is likely to be highly polygenic, FDR of 15% was used. Different distances (range 20kb-1MB) from lead SNP to closest gene (start and end site of longest transcript) were explored. All gene-sets were adjusted for gene-clustering: removing all but one gene from a subset of genes assigned the same best SNP and keeping the gene with the most significant gene score. Gene-set enrichment analyses were also attempted using three additional platforms: WebGestalt GSAT (www.webgestalt.org/), DAVID (https://david.ncifcrf.gov/) and Pascal (https://www2.unil.ch/cbg/index.php?title=Pascal). All three platforms yielded results identical to the ones obtained by MAGENTA.

5.3.5. Congenital cataract genes

I explored whether any genes previously linked to congenital cataract harboured age-related nuclear cataract association signals. The list of gene linked to congenital cataract was assembled using the following databases: OMIM — Online Mendelian Inheritance in Human, Cat-Map — Cataract Map; ClinVar — Clinical Variants, HD — in-house database assembled using extensive literature search and co-curated by Prof. Sudha Iyengar. In order to select genes whose involvement in congenital cataract pathogenesis was supported by multiple sources, only genes present in 2 or more of the database were considered. Genes with insufficient evidence for involvement were excluded. Such examples were genes supported by a single study only where a mutation present in single family lacked independent replication or any additional functional evidence within the report. In addition, genes with well-established involvement* in murine cataracts, irrespective of evidence from human studies, were also considered. In this manner, a final list of 96 genes was curated.

Next, I explored whether there was evidence for association between age-related nuclear cataract and variants mapping to the 96 genes within the discovery phase meta-analysis results. For a congenital cataract gene to be considered associated its

* multiple articles using a variety of molecular techniques

longest transcript had to contain or have in proximity multiple variants that were associated with age-related nuclear cataract (p-value < 0.05). Variants were assumed to be in proximity if they were located within a 100kb window around the gene or within a window demarcated by the two closest recombination hot-spots (recombination rates ≥ 40), whichever was smaller. For each gene, the p-value assigned to that gene was the p-value observed at the sentinel variant assigned to the gene.

5.3.6. Functional mapping of discover phase association signals

In order to predict the likely functional consequence of the associated variants and/or to predict causal effects, the sentinel SNPs (p-value < $1x10^{-5}$) in the discovery phase loci were taken forwards to bioinformatic analysis.

Regulatory elements: Variants were annotated in terms of evidence for regulatory functions (enhancer histone modification signals, DNase I hypersensitivity, binding of transcription factors or effects on regulatory motifs), using HaploReg[569] and ENCODE data track in the UCSC genome browser.Data on eye-specific transaction factors was also obtained using the Tissue Specific Gene Expression and Regulation database (TiGER)*. ENCODE (https://www.encodeproject.org/) aims to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which genes are active. HaploReg[569] intersects a variety of functional annotations beyond ENCODE data. Those come mostly from next-generation sequencing technologies and map DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts normal tissues[569]. TiGER is a database, developed by the Bioinformatics Lab at Wilmer Eye Institute, which contains tissue-specific gene expression profiles or expressed sequence tag (EST) data, cis-regulatory module data, and combinatorial gene regulation data (http://bioinfo.wilmer.jhu.edu/tiger/).

eQTL summary statistics: I explored whether any of the SNPs associated with nuclear cataract (p-value < $1x10^{-5}$), regulated gene expression of adjacent genes in a quantitative manner (eQTLs) by searching publicly available data (GTEx, http://www.gtexportal.org/home/) and the available literature. GTEx collects and analyses multiple human tissues from donors who are also densely genotyped, to assess genetic variation within their genomes. GTEx explores the global RNA expression

* http://bioinfo.wilmer.jhu.edu/tiger/

within individual tissues and treats the expression levels of genes as quantitative traits, thus identifying variations in gene expression that are highly correlated with genetic variation. For the purposes of the current study, eQTL from all available GTEx tissues were queried. In addition, I used eQTL summary-level data generated using the TwinsUK expression data[570,571] (N=856, lymphocytes, subcutaneous fat and skin) and genetic data imputed against 1000 Genome Phase 1 genotypes. The eQTL data was provided to me by Dr. Kerrin Small and Dr. Alexander Alves.

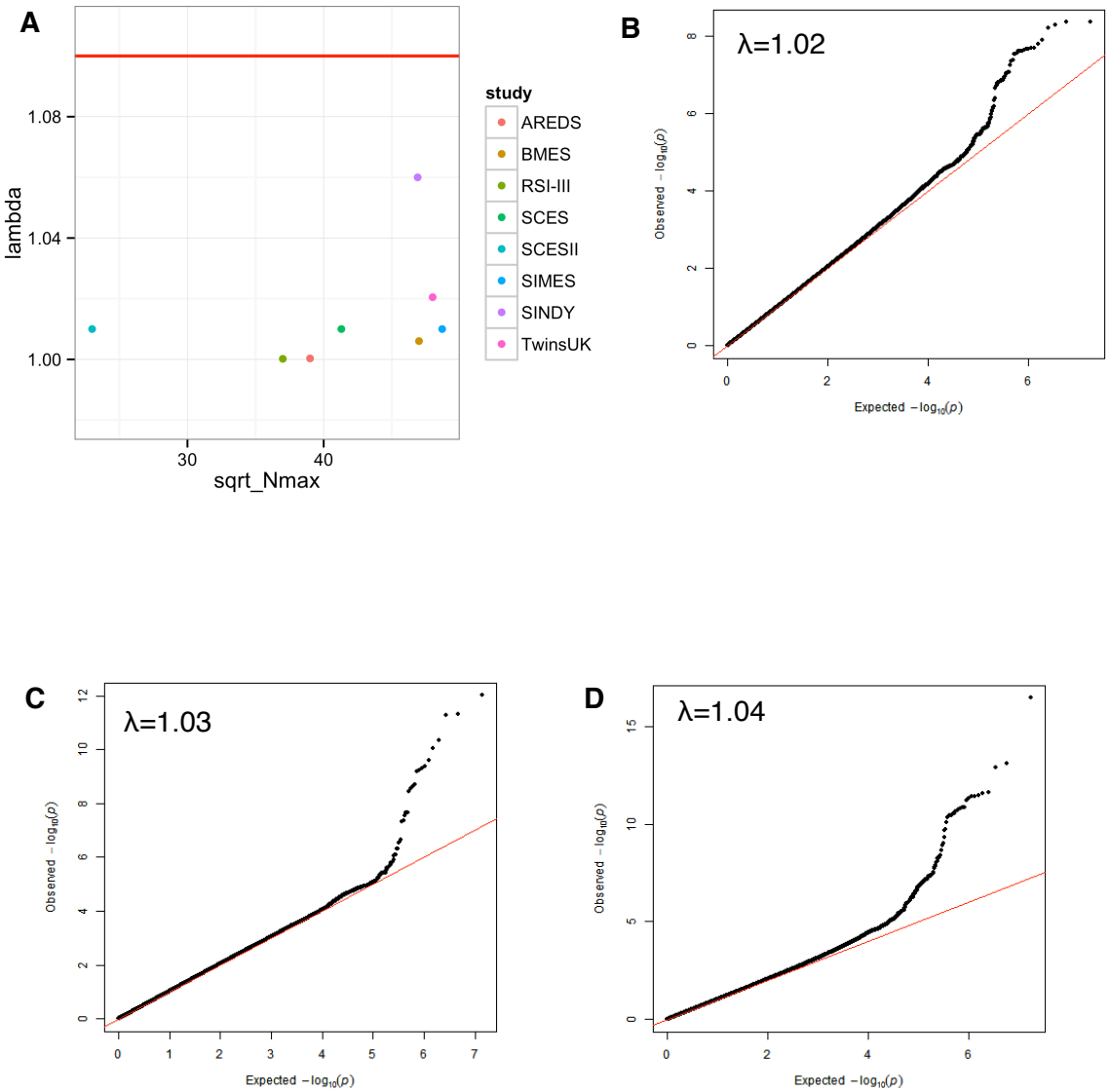
Gene Expression: Information on expression patterns in the tissue of interest can guide the selection of candidate genes in a locus of interest. As there is no in-house lens expression available, gene expression data was obtained from 4 publicly available databases for the gene closest to the most strongly associated variant per locus as well as for all other genes in LD (recombination rates ≥ 40). The databases used to obtain data gene expression in the human eye and lens were: iSyte[572] (https://research.bioinformatics.udel.edu/iSyTE/ppi/index.php), the Ocular Tissue Database (https://genome.uiowa.edu/otdb/), and TiGER. Gene expression data for mouse lens was obtained from Mouse Genome informatics (MGI, http://www.informatics.jax.org/). In addition, publicly available data on co-expression (CORD) was also queried[573]. The iSyte[572] aims to help candidate gene prioritisation relevant to lens development and cataract by proving information on transcriptome changes associated with cataract or lens defects, and by analysing gene expression in human and mouse embryos. The Ocular Tissue Database profiles RNA expression at a resolution of individual exons level from donor eye tissues using Affymetrix human Exon 1.0 ST arrays. MGI is an international database resource that provides integrated genetic, genomic, and biological data on laboratory mouse models in order to facilitate the study of human health and disease. Finally, CORD is a meta-analysis program that helps to determine gene function by exploring gene expression an co-expression, namely experiments where a gene is differentially expressed are identified and the gene network in which the gene belongs is determined[573]. COR also provided a dataset containing gene expression data from 9,490 microarray datasets mined from the GEO and ArrayExpress databases.

5.4. Results

After removing SNPs which were poorly imputed (info score/r^2 < 0.3), were rare (MAF < 1%) or out of HWE, 8.3 million variants in the European cohorts (N = 7,352) and 6.8 million variants in the Asian cohort (N = 6,799) were retained for analysis. The

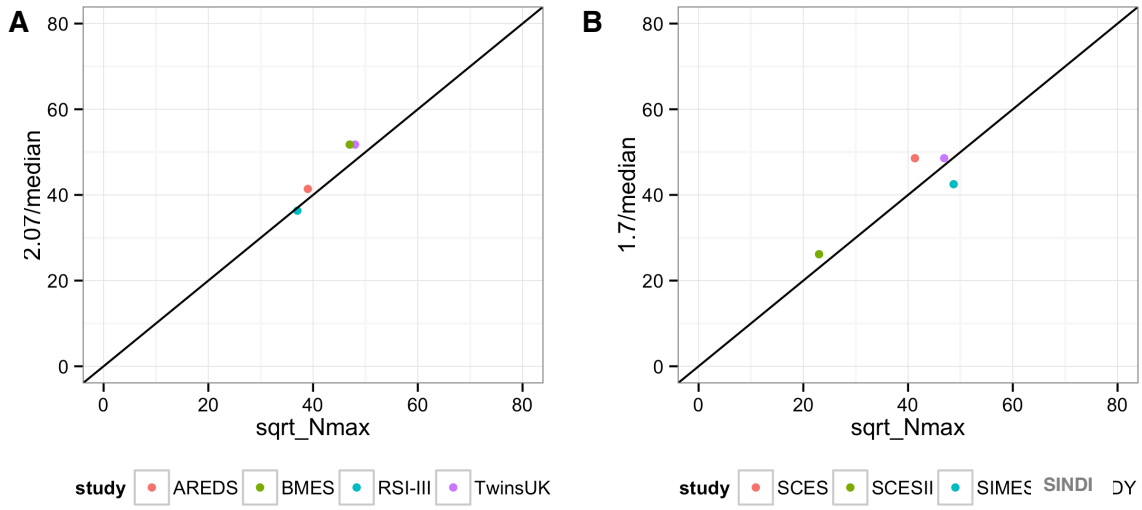
genomic inflation factors for the different studies were low (λ:0.99-1.04 **Figure 5.2.**) indicating excellent within-study control of population substructure. In the QC-ed subset of SNPs, the median standard errors (**Figures 5.3.**) were in accordance to the ones expected given the sample size of the different studies and the betas were normally distributed. The majority of associated loci showed low to moderate heterogeneity (**Figure 5.4**), therefore, fixed-effect meta-analysis were employed but validated those using a random-effects model.

Figure 5.2. Population stratification diagnostics



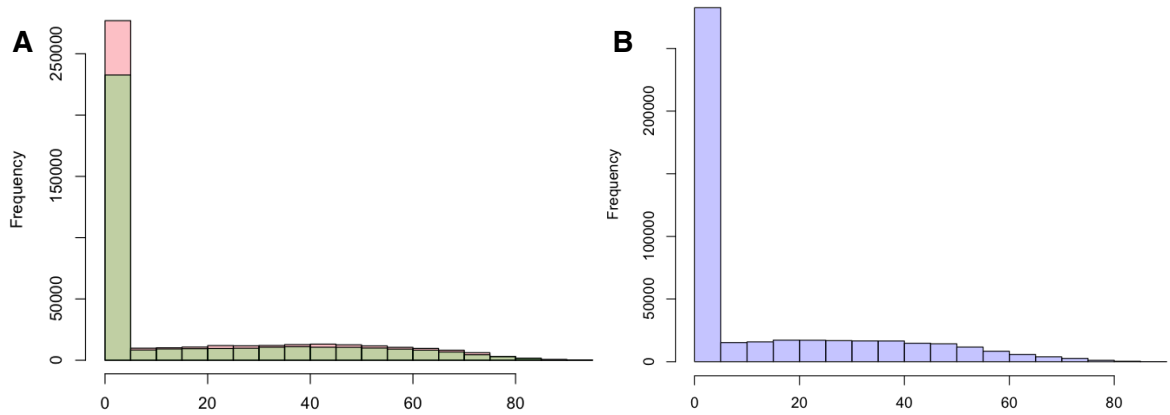
The Lambda-N plot (A) show the genomic inflation factor for each study plotted against the square root of the maximum sample size in that study; red line denotes threshold for inflation of λ=1.1. The meta-analyses Q-Q plots (B to D) show the observed distribution of p-values for association (-log10 transformed) plotted against the expected distribution of p-values under the null hypothesis in the European (B), Asian (C) and the combined (D) analysis.

Figure 5.3. SE-N diagnostic plots for trait distribution issues



This plot shows the results from the standard error (SE) versus sample-size (N) diagnostics analysis in European (A) and Asian (B) cohorts. X-axis: c/median SE where c is a constant calculated using the appropriate 1000G minor allele frequency data. Y-axis: the square root of the maximum sample size in each study.

Figure 5.4. Distribution of the meta-analysis heterogeneity (I^2) scores



This graph shows the frequency distribution plots of meta-analysis heterogeneity scores (I^2) of for variants (p-value < 0.05) associated with nuclear cataract in the European (pink) and Asian (green) ancestry cohorts (A) and the meta-analysis of all cohorts (B).

5.4.1. Discovery phase

Eighteen variants within the *SOX2-OT* locus and a variant within *LOC440704* reached genome-wide significance in the cohorts of European origin (**Figure 5.5.**). The most strongly associated variant at the *SOX2-OT* locus was rs9837662 (MAF=0.35; p-

value = $4.4x10^{-9}$) and it was also associated in Asians (p-value = $2.5x10^{-5}$). The variant within *LOC440704,* rs147580522 (MAF=2%, p-value = $4.4x10^{-8}$) appeared to be driven by the TwinsUK cohort (**Table 5.6.**). Moreover, it was absent in the Asian cohorts and in one of the two European replication cohorts. In the remaining BDES cohort, the variant did not replicate (p-value = 0.93, **Table 5.8.**). In addition, 118 variants were suggestively (p-value < $1x10^{-5}$) associated with nuclear cataract. These variants were within or in close proximity to 18 protein-coding genes, including *CRYAB* (rs10502150, p-value = $2.2x10^{-6}$) — a gene which encodes for an important structural lens protein. Moreover, *CRYAA* and *KCNAB1*, the two loci previously found to be associated with nuclear cataract in Asians[574], were also associated in Europeans (*CRYAA:* p-value = $5.3x10^{-6}$ at rs7278468; *KCNAB1:* p-value = 0.02 at rs55818638 (MAF=3%)). In Asian cohorts (N=6,799) nineteen variants in *CRYAA* (rs7278468, p-value = $9.8x10^{-13}$) and *GJA3* (rs4769087, p-value = $2.2x10^{-8}$) loci, reached genome-wide significance (**Figure 5.5.**). In addition, 59 variants within 24 loci were suggestively associated with nuclear cataract (p-value < $1x10^{-5}$), including SNPs within *KCNAB1* (rs7615568, p-value = $3.9x10^{-6}$).

When the European and Asian ancestry cohorts were combined ($N_{variants}$=8.5 million), SNPs at three loci reached genome-wide significance (**Figure 5.5.**). Two of the loci, *CRYAA* (rs7278468, p-value = $3.6x10^{-17}$) and *SOX2-OT* (rs9842371, p-value = $2.6x10^{-12}$) were the most strongly associated in each ethnic group, while variants within *TMPRSS5* (rs4936279, p-value = $4.2x10^{-11}$) reached genome-wide significance only after combining the two ancestry groups. In addition, rs62149908 (within *COMMD1*) fell short of reaching genome-wide significance (p-value = $6.5x10^{-8}$). Moreover, 30 loci were suggestively associated with nuclear cataract (p-value < $1x10^{-5}$, **Table 5.3**), including *GJA3* (p-value = $6.6x10^{-6}$) and *CRYAB* (p-value = $8.4x10^{-6}$). The gene-based analysis (**Table 5.4**) did not yield additional associated genes. The common variants we found associated with age-related nuclear cataract (p-value < $1x10^{-5}$) explained 3% of genetic variance.

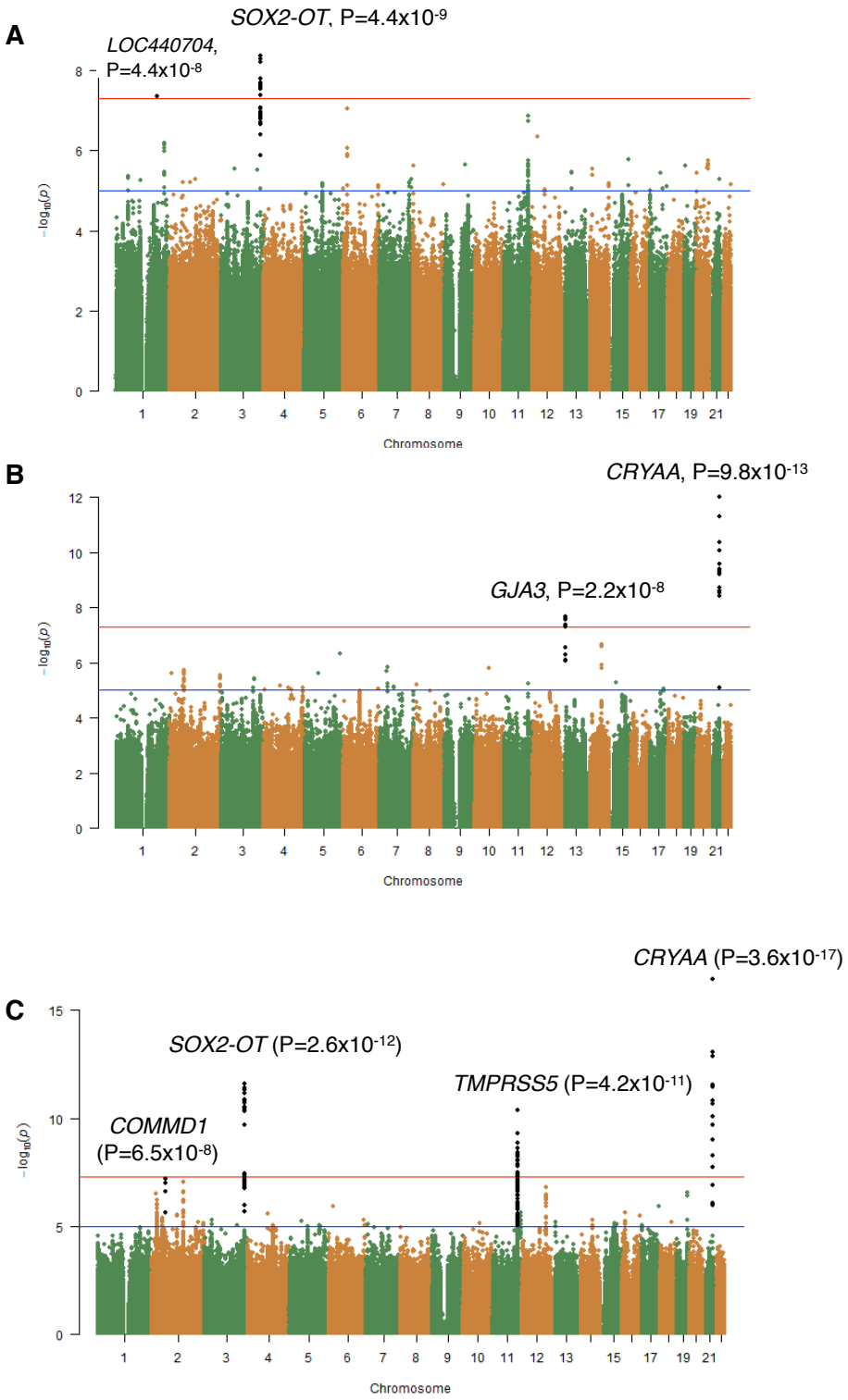
Table 5.3: Discovery phase meta-analysis results

Position						Combined analysis				Europeans				Asians			
CHR	BP	SNP	Gene	EA/OA	EAF (EUR/AS)	Beta	SE	p-value	I^2	Beta	SE	p-value	I^2	Beta	SE	p-value	I^2
1q31.1	**190717899**	**rs147580522**	***LOC440704*** [b]	**C/G**	**0.02/0.00**	**0.48**	**0.09**	**4.4E-08**	**38.4**	**0.48**	**0.09**	**4.4E-08**	**38.4**	**-**	**-**	**-**	**-**
2p24.1	**20747778**	**rs61185326**	**intergenic**	**T/A**	**0.04/0.02**	**0.20**	**0.04**	**3.0E-07**	**50.1**	**0.18**	**0.04**	**1.3E-05**	**46.9**	**-**	**-**	**-**	**-**
2p23.3	**24439276**	**rs13021828**	***ITSN2*** [b]	**C/G**	**0.38/0.39**	**-0.05**	**0.01**	**6.1E-07**	**23.6**	**-0.05**	**0.02**	**1.4E-03**	**57.2**	**-0.05**	**0.01**	**1.3E-04**	**0.0**
2p15	**62191878**	**rs62149908**	***COMMD1*** [b]	**T/C**	**0.22/0.32**	**-0.06**	**0.04**	**6.5E-08**	**53.5**	**-0.06**	**0.09**	**2.5E-03**	**77.5**	**-0.06**	**0.04**	**7.3E-06**	**0.0**
2q14.1	115778290	rs77791858	*DPP10* [b]	T/C	0.03/0.02	0.19	0.04	3.1E-06	28.4	0.16	0.05	6.0E-04	0.0	-	-	-	-
2q22.3	**145341259**	**rs16823886**	***LINC01412*** [d]	**A/G**	**0.13/0.18**	**-0.06**	**0.01**	**8.8E-08**	**57.4**	**-0.08**	**0.02**	**3.2E-04**	**78.6**	**-0.06**	**0.01**	**6.0E-05**	**0.0**
2q34	211624516	rs36191905	intergenic	C/T	0.48/0.49	0.05	0.01	6.2E-06	10.6	0.04	0.02	1.4E-02	0.0	0.05	0.01	1.1E-04	39.9
2q37.1	235342061	rs10199789	intergenic	C/T	0.21/0.11	0.06	0.01	7.5E-06	53.0	0.03	0.02	7.8E-02	36.0	0.09	0.02	3.0E-06	43.9
3p24.1	30813873	rs1827496	*GADL1* [b]	A/G	0.32/0.15	0.05	0.01	5.2E-06	0.0	0.06	0.02	2.7E-04	0.0	0.05	0.02	5.3E-03	0.0
3q24	148672923	rs76405079	intergenic	C/T	0.09/0.04	-0.09	0.02	9.1E-06	52.1	-0.07	0.03	6.3E-03	73.1	-0.12	0.03	2.1E-04	0.0
3q26.33	**181346937**	**rs9842371**	***SOX2-OT*** [b]	**G/A**	**0.35/0.54**	**0.07**	**0.01**	**2.6E-12**	**40.9**	**0.07**	**0.01**	**2.6E-12**	**40.9**	**0.05**	**0.01**	**1.5E-05**	**0.0**
4q22.1	91932333	rs72667512	*CCSER1* [b]	G/A	0.41/0.17	0.06	0.01	2.6E-06	0.0	0.06	0.02	1.7E-04	0.0	0.05	0.02	4.7E-03	0.0
4q28.1	127728164	rs77573620	intergenic	T/A	0.02/0.07	0.17	0.04	6.5E-06	0.0	-	-	-	-	0.16	0.04	2.1E-05	0.0
5q11.22	55870395	rs71624138	*C5orf67* [b]	A/G	0.12/0.06	-0.08	0.02	5.5E-06	0.0	-0.10	0.02	1.4E-05	0.0	-0.05	0.03	5.8E-02	0.0
5q31.3	139543125	rs156084	*LOC101929719* [b]	A/G	0.27/0.57	0.04	0.01	9.1E-06	0.0	0.03	0.02	5.6E-02	0.0	0.05	0.01	4.3E-05	0.0
5q31.3	142111021	rs72796639	intergenic	G/A	0.03/0.02	0.14	0.03	1.0E-05	0.0	0.17	0.04	1.8E-05	0.0	0.08	0.05	7.9E-02	0.0
6q25.3	159308438	rs34891821	*C6orf99* [b]	G/T	0.32/0.12	0.07	0.02	5.1E-06	9.1	0.07	0.02	2.4E-05	31.7	-	-	-	-
6q27	165953127	rs545099	*PDE10A* [b]	C/T	0.02/0.04	0.16	0.04	9.3E-06	0.0	0.09	0.06	1.3E-01	0.0	0.20	0.04	9.2E-06	0.0
10q22.1	73199588	rs41281302	*CDH23* [a]	T/T	0.03/0.04	0.17	0.04	6.9E-06	0.0	0.15	0.05	2.1E-03	0.0	-	-	-	-
11q23.1	111788464	rs10789852	HSPB2 [b]/C11orf52 (*CRYAB*)	G/A	0.35/0.49	-0.04	0.01	8.4E-06	0.0	-0.07	0.02	2.5E-05	0.0	-0.03	0.01	1.9E-02	0.0
11q23.2	**113566207**	**rs4936279**	***TMPRSS5*** [b]	**A/C**	**0.30/0.48**	**0.06**	**0.01**	**4.2E-11**	**10.8**	**0.08**	**0.02**	**1.3E-07**	**6.3**	**0.05**	**0.01**	**1.8E-05**	**0.0**
11q24.2	126058369	rs487450	intgenic	A/C	0.24/0.03	-0.07	0.01	2.4E-06	19.2	-0.06	0.02	2.5E-04	46.5	-0.08	0.03	2.7E-03	0.0

11q24.3	130270855	rs59317921	*ADAMTS8* [d]	T/C	0.85/0.09	0.06	0.01	3.4E-06	0.0	0.06	0.02	4.4E-05	0.0	0.05	0.02	2.2E-02	0.0
12q24.11	**109988214**	**rs11067211**	***MMAB*** [d]	**G/A**	**0.26/0.16**	**0.06**	**0.01**	**1.6E-07**	**0.0**	**0.06**	**0.02**	**3.9E-04**	**0.0**	**0.06**	**0.02**	**1.1E-04**	**0.0**
13q12.11	20721897	rs17077135	*GJA3* [b]	C/T	0.11/0.40	0.05	0.01	6.2E-06	68.8	0.02	0.02	3.3E-01	58.0	0.07	0.01	2.7E-08	26.2
14q24.2	73277092	rs1990441	*DPF3* [b]	C/A	0.47/0.49	0.05	0.01	5.4E-06	0.0	0.02	0.02	2.8E-01	0.0	0.06	0.01	1.6E-06	0.0
15q23	69096596	rs11639228	*ANP32A* [b]	T/G	0.42/0.31	0.04	0.01	7.2E-06	1.1	0.05	0.01	4.1E-04	34.6	0.04	0.01	4.0E-03	0.0
15q24.2	75449755	rs143061779	intergenic	T/C	0.12/0.40	-0.06	0.01	8.6E-06	0.0	-0.05	0.03	7.6E-02	0.0	-0.06	0.01	4.0E-05	0.0
16p13.11	15807137	rs12324990	*NDE* [b] */MIH11*	A/C	0.04/0.03	0.12	0.03	2.4E-06	0.0	0.10	0.04	7.7E-03	0.0	0.14	0.03	7.8E-05	0.0
16q23.2	79468817	rs28703582	intergenic	G/A	0.50/0.41	-0.05	0.01	3.1E-06	15.1	-0.04	0.02	4.3E-03	19.2	-0.05	0.01	2.2E-04	33.4
17p13.3	1654089	rs56679346	*SERPINF2* [b]	G/C	0.28/0.45	0.05	0.01	8.8E-06	53.2	0.08	0.02	9.9E-06	0.0	0.03	0.01	3.6E-02	69.5
17q25.3	76170735	rs1065769	*TK1* [a]	T/C	0.32/0.18	0.05	0.01	1.2E-06	0.0	0.06	0.02	2.5E-04	0.0	0.05	0.02	1.3E-03	0.0
18q21.31	55659197	rs1942539	intergenic	T/G	0.26/0.29	-0.05	0.01	6.5E-06	0.0	-0.05	0.02	5.4E-03	0.0	-0.05	0.02	3.8E-04	0.0
19q13.33	**48206092**	**rs1005911**	***GLTSCR1*** [a]	**G/T**	**0.25/0.36**	**-0.05**	**0.01**	**2.8E-07**	**0.0**	**-0.06**	**0.02**	**7.1E-04**	**0.0**	**-0.05**	**0.01**	**1.0E-04**	**0.0**
21q22.3	**44588757**	**rs7278468**	***CRYAA*** [c]	**G/T**	**0.69/0.40**	**0.08**	**0.01**	**3.6E-17**	**20.5**	**0.07**	**0.02**	**5.3E-06**	**0.0**	**0.09**	**0.01**	**9.8E-13**	**59.5**

This table presents the discovery phase meta-analysis results. The sentinel SNPs (dbSNP147 rs-number) at each locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; EA – effect allele; OA – other allele; EAF – effect allele frequency in the European (EUR) and Asian (AS) ancestry cohorts; The results for association between standardised nuclear cataract scores and genetic variants, using inverse-variance fixed effect meta-analysis model are presented as follows: Beta — effect size; SE — standard error of the beta, p-value and heterogeneity score (I^2) are listed for the combined analysis and for the analysis within each ancestry group.Variants in bold (p-value < 10^{-6}) were taken forwards for replication.

Figure 5.5. Discovery phase Manhattan plots



This figure presents Manhattan plots for the European ancestry (A), Asian ancestry (B) and combined (C) meta-analyses: the -log10 transformed p-values for association are plotted against the 22 autosomes. The red line denotes the genome-wide significance threshold (p-value = $5x10^{-8}$); the blue line denotes the suggestive significance threshold (p-value = $1x10^{-5}$).

Table 5.4: Gene-based analysis of the Discovery phase loci

Gene	p-value	corrected p-value
CRYAA	2.3E-15	2.9E-11
SOX2-OT	1.2E-09	9.6E-06
TMPRSS5	2.2E-09	1.4E-05
LINC01412	5.4E-06	0.02
LOC440704	5.7E-06	0.02
MMAB	7.1E-06	0.02
COMMD1	7.7E-06	0.02
STAB2	1.7E-05	0.04
ITSN2	1.8E-05	0.04

The table shows the results from the gene-based analysis (GATES) of the summary statistics from the Discovery phase: combined analysis of European and Asian ancestry cohorts. Both the raw p-value and the p-value corrected for number of genes tested are presented.

5.4.2. Validation, trans-ethnic and conditional analyses

In order to ensure that by selecting a fixed-effect model no loci were missed, random-effect meta-analysis were also performed (**Table 5.5.**). The random-effect model produced similar results to the fixed effect model and no additional loci were identified. Next, leave-one-out sensitivity analyses were performed to evaluate the robustness of the association results (**Table 5.6**). The effect size and direction of effect stayed stable, thus the results did not support a disproportionate influence of any of the participating cohorts. The only exception was the sentinel variant at the 2p24.1 where BMES strongly influenced the association. This locus however did not replicate (**Table 5.8**).

To further evaluate the sharing of genetic factors between the European and Asian cohorts, trans-ethnic meta-analysis were performed **(Table 5.7)** in two different ways, both of which use a Bayesian framework but employ different methods to calculate the population relatedness. The first method, implemented in MANTRA[566], assigns studies to population clusters and the effect of each study is compared to the cluster effect. The distance between the cluster centre and each study is calculated on the basis of allele frequency dissimilarity. Due to the fact that this method is computationally intensive and data formatting is cumbersome, only a subset of the SNPs (p-value < $1x10^{-4}$) were used to calculate the distance between clusters. The second method, implemented in MR-MEGA*, calculates distance matrixes using the

* http://www.geenivaramu.ee/en/tools/mr-mega

mean pairwise differences between effect allele frequencies across a set of thinned SNPs. The matrices are then subjected to multidimensional scaling in order to calculate principal components, which are in turn used as covariates in a linear inverse-variance weighted model. In addition, the heterogeneity between studies is partitioned into ancestry-related heterogeneity and residual heterogeneity.

The results from the trans-ethnic meta-analysis **(Table 5.7)** again reinforced the observation that, with two exceptions (the *GJA3* and *CRYAB* loci), the observed genetic effects on nuclear cataract were ancestry-independent; and no additional loci reached genome-wide significance. The LD-structure of the *GJA3* locus differs between European and Asian ancestry populations (**Figure 5.6**). Moreover, the posterior betas calculated by MANTRA had opposite direction of effect in the two populations. There was some evidence for ancestry-independent heterogeneity in this locus, too. There were no differences in LD-structure for the *CRYAB* locus between the two ancestry groups. Due to allele frequencies, however, for every associated variant the effect was much stronger in Europeans as opposed to Asians. The *COMMD1* locus showed statistically significant heterogeneity that could not be explained by ancestry effects. At this locus the source of heterogeneity was the TwinsUK cohort which showed opposite direction of effect to the rest of the cohorts. The result was not due to genotyping errors, imputation errors or outlier effects. Interestingly, RNA sequencing analysis in human lenses with or without nuclear cataract performed by Dr. Chaolong Wang (data not shown), suggested *COMMD1* expression differences between females and males pointing to an opposite direction of effect in females. Given that 98% of TwinsUK participants are female the result may reflect a genuine biological effect.

Table 5.5: Comparison between the fixed-effect and the random-effect models

Position						Fixed Effect Meta-analysis				Random Effect meta-analysis			
CHR	BP	SNP	Gene	EA/OA	EAF (EUR/AS)	Beta	SE	p-value	I^2	Beta	SE	RE p-value	EH p-value
1q31.1	190717899	rs147580522	*LOC440704* [b]	C/G	0.02/0.00	0.48	0.09	4.4E-08	38.4	0.48	0.09	5.1E-04	6.5E-08
2p24.1	20747778	rs61185326	intergenic	T/A	0.04/0.02	0.20	0.04	3.0E-07	50.1	0.22	0.06	2.2E-04	2.6E-07
2p23.3	24439276	rs13021828	*ITSN2* [b]	C/G	0.38/0.39	-0.05	0.01	6.1E-07	23.6	-0.05	0.01	1.6E-05	1.1E-06
2p15	62191878	rs62149908	*COMMD1* [b]	T/C	0.22/0.32	-0.06	0.04	6.5E-08	53.5	-0.06	0.02	7.5E-04	5.2E-08
2q22.3	145341259	rs16823886	*LINC01412* [d]	A/G	0.13/0.18	-0.06	0.01	8.8E-08	57.4	-0.07	0.02	4.6E-04	5.5E-08
3q26.33	181346937	rs9842371	*SOX2-OT* [b]	G/A	0.35/0.54	0.07	0.01	2.6E-12	40.9	0.07	0.01	3.4E-08	4.8E-12
11q23.2	113566207	rs4936279	*TMPRSS5* [b]	A/C	0.30/0.48	0.06	0.01	4.2E-11	10.8	0.06	0.01	4.5E-10	8.3E-11
12q24.11	109988214	rs11067211	*MMAB* [d]	G/A	0.26/0.16	0.06	0.01	1.6E-07	0.78	0.06	0.01	1.6E-07	2.9E-07
19q13.33	48206092	rs1005911	*GLTSCR1* [a]	G/T	0.25/0.36	-0.05	0.01	2.8E-07	0.0	-0.05	0.01	2.8E-07	5.1E-07
21q22.3	44588757	rs7278468	*CRYAA* [c]	G/T	0.69/0.40	0.08	0.01	3.6E-17	20.5	0.08	0.01	2.3E-13	7.2E-17

The comparison results between the fixed-effect and the random-effect models for the loci which were selected for replication are presented here. The sentinel SNPs (dbSNP147 rs-number) at each locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; EA – effect allele; OA – other allele; EAF – effect allele frequency in the European (EUR) and Asian (AS) ancestry cohorts; The results for association between standardised nuclear cataract scores and genetic variants (combined analysis), presented as follows: Inverse-variance fixed-effect meta-analysis model estimates: Beta — effect size; SE — standard error of the beta, p-value and heterogeneity score (I^2) are listed. for the combined analysis and for the analysis within each ancestry group.Variants in bold (p-value < 10^{-6}) were taken forwards for replication. Random-effect meta-analysis estimates: Beta — effect size; SE — standard error of the beta, p-value (RE) calculated using conventional inverse-variance random effect model and p-value (EH) using Han and Eskin's random effects model are presented.

Table 5.6. Sensitivity analyses

Position					All studies		AREDS		BMES		TwinsUK		RSI-III		SIMES		SINDI		SCESI		SCESII	
CHR	BP	SNP	Gene	EA/ OA	Beta	SE	Beta	SE	Beta	SE	Beta	SE	Beta	SE	Beta	SE	Beta	SE	Beta	SE	Beta	SE
1q31.1	190717899	rs147580522	*LOC440704* [b]	C/G	0.48	0.09	0.51	0.11	-	-	0.35	0.12	0.54	0.10	-	-	-	-	-	-	-	-
2p24.1	20747778	rs61185326	intergenic	T/A	0.20	0.04	0.21	0.04	0.16	0.05	0.28	0.05	0.20	0.04	-	-	0.18	0.04	-	-	-	-
2p23.3	24439276	rs13021828	*ITSN2* [b]	C/G	-0.05	0.01	-0.05	0.01	-0.05	0.01	-0.06	0.01	-0.05	0.01	-0.05	0.01	-0.05	0.01	-0.05	0.01	-0.06	0.01
2p15	62191878	rs62149908	*COMMD1* [b]	T/C	-0.06	0.04	-0.06	0.02	-0.05	0.01	-0.07	0.01	-0.06	0.01	-0.06	0.01	-0.06	0.02	-0.06	0.01	-0.06	0.01
2q22.3	145341259	rs16823886	*LINC01412* [d]	A/G	-0.06	0.01	-0.06	0.01	-0.06	0.01	-0.07	0.01	-0.06	0.01	-0.07	0.01	-0.07	0.01	-0.07	0.01	-0.06	0.01
3q26.33	181346937	rs9842371	*SOX2-OT* [b]	G/A	0.07	0.01	0.06	0.01	0.07	0.01	0.06	0.01	0.07	0.01	0.07	0.01	0.07	0.01	0.07	0.01	0.07	0.01
11q23.2	113566207	rs4936279	*TMPRSS5* [b]	A/C	0.06	0.01	0.06	0.01	0.01	0.07	0.07	0.01	0.06	0.01	0.07	0.01	0.07	0.01	0.06	0.01	0.06	0.01
12q24.11	109988214	rs11067211	*MMAB* [d]	G/A	0.06	0.01	0.06	0.02	0.07	0.01	0.06	0.01	0.06	0.01	0.06	0.01	0.06	0.01	0.06	0.01	0.06	0.01
19q13.33	48206092	rs1005911	*GLTSCR1* [a]	G/T	-0.05	0.01	-0.05	0.01	-0.05	0.01	-0.06	0.01	-0.05	0.01	-0.06	0.01	-0.06	0.01	-0.05	0.01	-0.05	0.01
21q22.3	44588757	rs7278468	*CRYAA* [c]	G/T	0.08	0.01	0.09	0.01	0.08	0.01	0.08	0.01	0.08	0.01	0.10	0.01	0.09	0.01	0.08	0.01	0.08	0.01

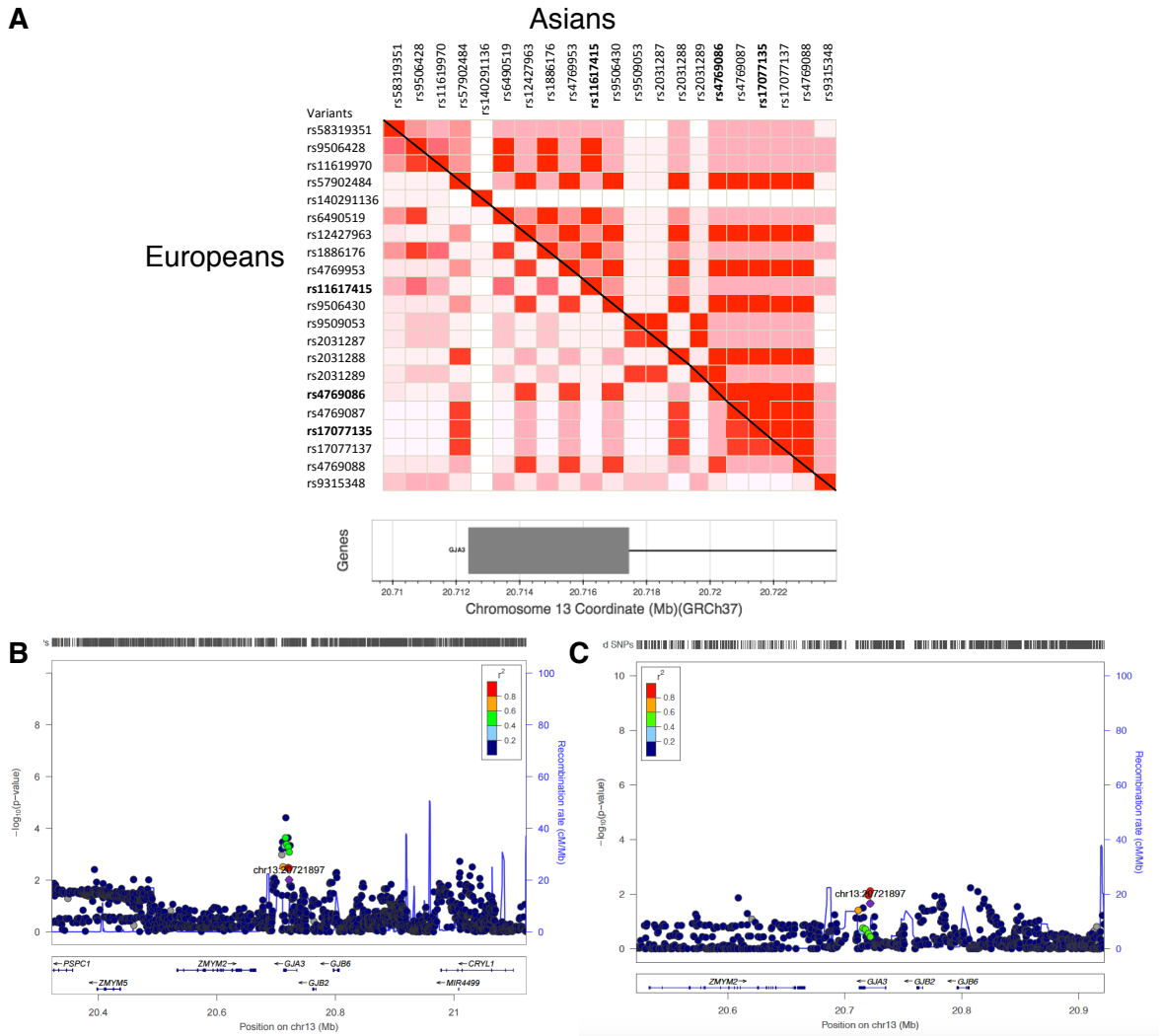
This table exemplifies the sensitivity analysis undertaken in the present study. The results for the loci taken forwards to replication are presented. Only the sentinel variant (dbSNP147 rs-number) is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; EA – effect allele; OA – other allele; The combine discovery phase beta and and its standard error (SE) are presented (All studies) follow by the analysis where one study at a time is dropped out. The name of the study that was excluded at each step is listed above the resulting beta and SE. AREDS – Age-related Eye Diseases Study (and its grading system), BMES – Blue Mountains Eye Study, RSI-III – Rotterdam Study I (follow-up 3), SCES – Singapore Chinese Eye Study; SiMES – Singapore Malay Eye Study; SINDI – Singapore Indian Eye Study.

Table 5.7: Results of the trans-ethnic meta-analyses

Position							MANTRA		MR-MEGA			
CHR	BP	SNP	Gene	EA/OA	EAF (EUR/AS)	FM p-value	BF	PP	BF	p-value	ANS p-value	HET p-value
1q31.1$	190717899	rs147580522	*LOC440704* b)	C/G	0.02/0.00	4.4E-08	-	-	-	-	-	-
2p24.1	20747778	rs61185326	intergenic	T/A	0.04/0.02	3.0E-07	5.4	0.40	13.7	4.51E-07	0.05	0.38
2p23.3	24439276	rs13021828	*ITSN2* b)	C/G	0.38/0.39	6.1E-07	5.0	0.06	10.4	5.63E-06	0.33	0.22
2p15	62191878	rs62149908	*COMMD1* b)	T/C	0.22/0.32	6.5E-08	5.8	0.11	12.4	8.27E-07	0.40	**0.02**
2q22.3	145341259	rs16823886	*LINC01412* d)	A/G	0.13/0.18	8.8E-08	5.8	0.13	11.2	2.58E-06	0.10	**0.006**
3q26.33	181346937	rs9842371	*SOX2-OT* b)	G/A	0.35/0.54	2.6E-12	10.2	0.14	24.2	8.17E-12	0.06	0.29
11q23.1	111788464	rs10789852	HSPB2 b)/C11orf52 (*CRYAB*)	G/A	0.35/0.49	8.4E-06	3.0	0.25	9.8	4.95E-05	**0.04**	0.63
11q23.2	113566207	rs4936279	*TMPRSS5* b)	A/C	0.30/0.48	4.2E-11	9.0	0.08	20.02	4.93E-10	0.25	0.41
12q24.11	109988214	rs11067211	*MMAB* d)	G/A	0.26/0.16	1.6E-07	5.5	0.05	11.4	2.28E-06	0.48	0.74
13q12.11	20721897	rs17077135	*GJA3* b)	C/T	0.11/0.40	6.2E-06	4.4	0.68	11.7	1.60E-06	**0.010**	**0.02**
19q13.33	48206092	rs1005911	*GLTSCR1* a)	G/T	0.25/0.36	2.8E-07	5.3	0.07	10.1	7.76E-06	0.97	0.64
21q22.3	44588757	rs7278468	*CRYAA* c)	G/T	0.69/0.40	3.6E-17	15.2	0.26	34.4	3.63E-16	0.14	0.44

This table presents the results from the trans-ethnic meta-analysis. $ Given that rs147580522 was absent from the Asian ancestry cohorts, trans-ethnic meta analysis for this variant were not possible. Only the sentinel variant (dbSNP147 rs-number) per locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene a) exonic/UTR; b) intronic; c) upstream; d) downstream; EA – effect allele; OA – other allele; EAF – effect allele frequency in the European (EUR) and Asian (AS ancestry cohorts; FM p-value: the p-value from the fixed-effect inverse variance meta-analysis; BF – Bayesian factor; posterior probability (PP) for heterogeneity; p values for association; ANS p-value – p-value for heterogeneity due to ancestry; HET p-value – p-value for residual heterogeneity.

Figure 5.6. Ethnic differences at the *GJA3* locus



This graphs shows the differences in LD-structure (A) between Europeans and Asians for the *GJA3* locus, and regional association plots where instead of plotting the p-value for association, the p-values for heterogeneity due to ancestry (B) and residual heterogeneity (C) are plotted.

5.4.3. Replication and all data meta-analysis

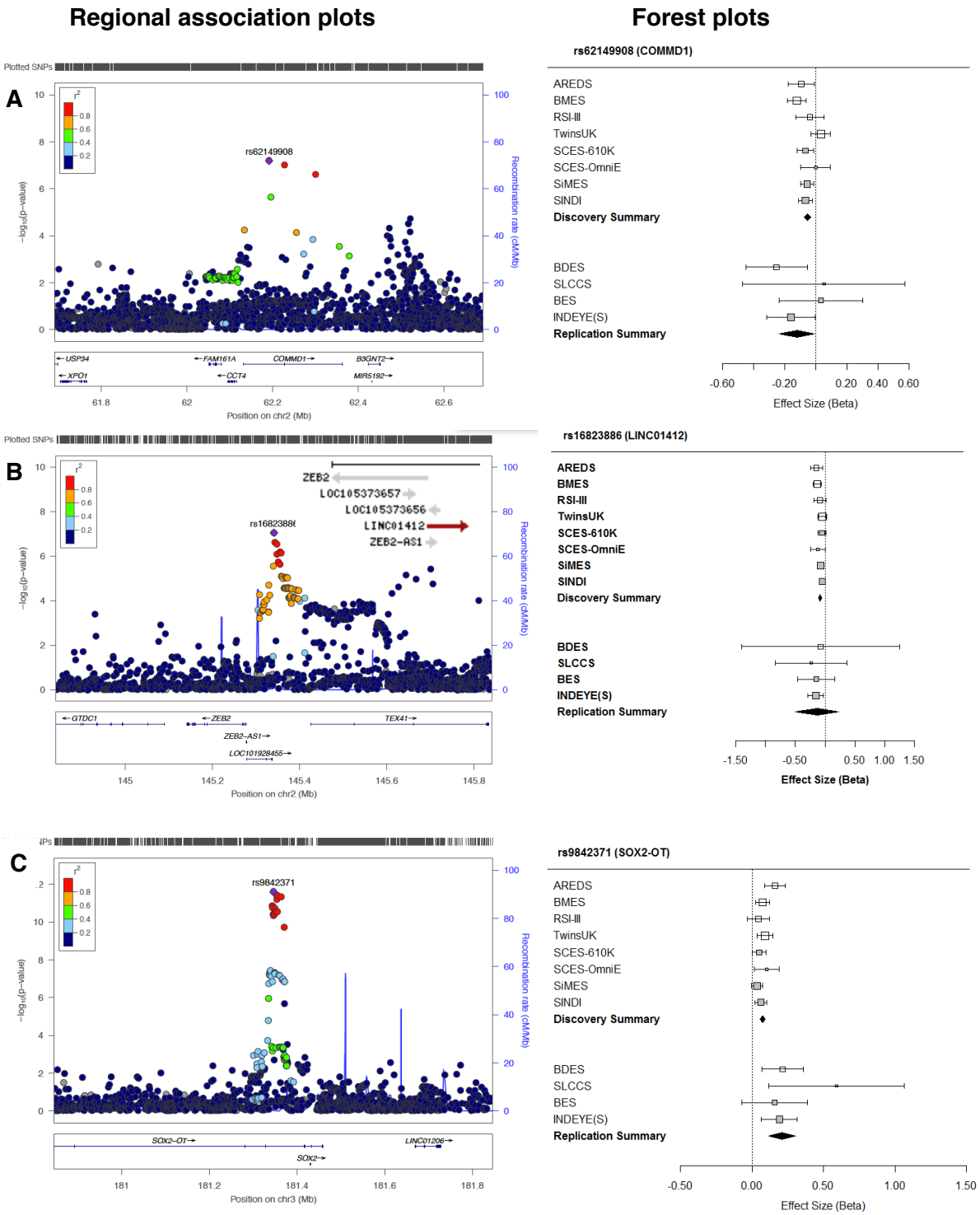
The most strongly associated variant per locus at ten discovery phase loci (p-value < 10^{-6}) were selected for replication. Four of the variants replicated (p-value < 0.05) and three SNPs, although not reaching nominal significance, had the same direction of effect as that in the discovery phase (**Table 5.8.**). Finally, the discovery and replication phase results were combined (all data meta-analysis) using sample-size-weighted p-value of the different studies (**Table 5.8.**). This resulted in total of 7 loci reaching genome-wide significance. Regional association and forest plots for those loci are presented in **Figure 5.7.**

Table 5.8. Replication and all-data meta-analyses

Position					Discovery phase (N=14151)					Replication phase (N=1664 cases /3961 controls)					All Data (N=19826)	
CHR	BP	SNP	Gene	EA/ OA	EAF (EUR/AS)	Beta	SE	p-value	I^2	EAF (cases/controls)	OR	SE	p-value	I^2	Z	p-value
1q31.1	190717899	rs147580522	*LOC440704* [b)]	C/G	0.02/0.00	0.48	0.09	4.4E-08	38.4	0.02/0.02*	0.95	0.55	0.93	-	4.79	1.6E-06
2p24.1	20747778	rs61185326	intergenic	T/A	0.04/0.02	0.20	0.04	3.0E-07	50.1	0.03/0.04	0.55	0.24	0.02	0.28	2.08	0.04
2p23.3	24439276	rs13021828	*ITSN2* [b)]	C/G	0.38/0.39	-0.05	0.01	6.1E-07	23.6	0.41/0.39	1.98	0.05	0.21	54.9	-3.03	2.4E-03
2p15	62191878	rs62149908 [$]	*COMMD1* [b)]	T/C	0.22/0.32	-0.06	0.04	6.5E-08	53.5	0.26/0.30	0.88	0.24	0.04	0.0	-5.8	5.6E-09
2q22.3	145341259	rs16823886	*LINC01412* [d)]	A/G	0.13/0.18	-0.06	0.01	8.8E-08	57.4	0.17/0.22	0.86	0.06	0.01	0.0	-6.0	1.7E-09
3q26.33	181346937	rs9842371	*SOX2-OT* [b)]	G/A	0.35/0.54	0.07	0.01	2.6E-12	40.9	0.40/0.35	1.23	0.05	3.4E-05	0.0	9.2	4.8E-20
11q23.2	113566207	rs4936279 [$]	*TMPRSS5* [b)]	A/C	0.30/0.48	0.06	0.01	4.2E-11	10.8	0.38/0.32	1.08	0.05	0.13	37.5	6.8	7.7E-12
12q24.11	109988214	rs11067211 [$]	*MMAB* [d)]	G/A	0.26/0.16	0.06	0.01	1.6E-07	0.0	0.22/0.21	1.15	0.06	0.06	0.0	5.5	3.8E-08
19q13.33	48206092	rs1005911	*GLTSCR1* [a)]	G/T	0.25/0.36	-0.05	0.01	2.8E-07	0.0	0.28/0.30	0.85	0.06	0.01	0.0	-5.7	1.2E-08
21q22.3	44588757	rs7278468 [$]	*CRYAA* [c)]	G/T	0.69/0.40	0.08	0.01	3.6E-17	20.5	0.47/0.36	1.12	0.07	0.17	36.8	7.8	6.8E-15

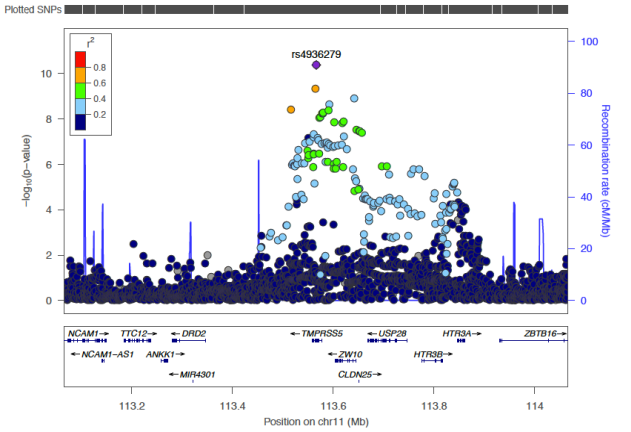
List of the SNPs that reached genome-wide significance ($P<5\times10^{-8}$, chr.:base pair, closest gene). Effect allele (EA); EA frequency (F) in % per stage or [a] in cases / controls. Stage (S): European ancestry (1), Asian ancestry (2), and All discovery studies (3); effect size (beta) and its standard error (SE); p-value (P) for association; heterogeneity values (I^2); p-value for heterogeneity (P_{het}); odds ratio (OR) from the replication case-control collections; Z-score (Z) derived from the all data meta-analysis. *in BDES only [b] SNPs not available in the INDEYE(S) study due to probe design issues. The following variants in high LD with the lead SNP were genotyped instead: rs55785307 (*COMMD1*, R^2=0.84; D'= 0.99); rs11601037 (*TMPRSS5*, R^2=0.90; D'=1.0), rs7486178 (*MMAB*, R^2=0.83; D'=0.99), and rs870137 (*CRYAA*, R^2=0.48; D'= 0.98).

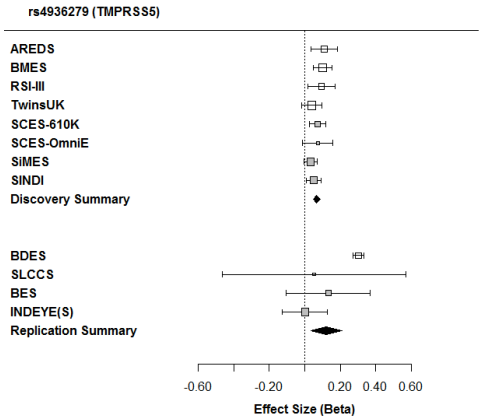
Figure 5.7. Regional association and forest plots for the loci associated with age-related nuclear cataract at the All Data meta-analysis phase



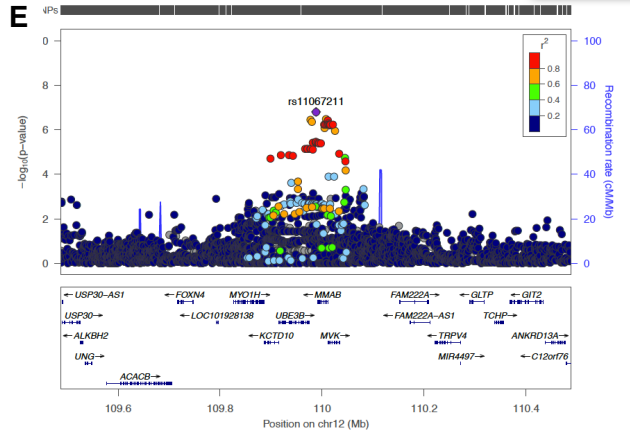
Regional association plots

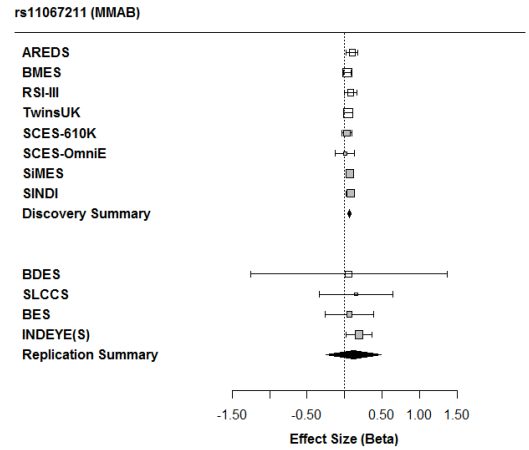
Forest plots



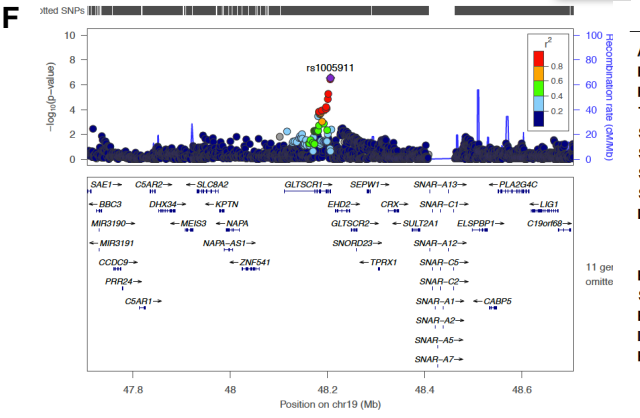


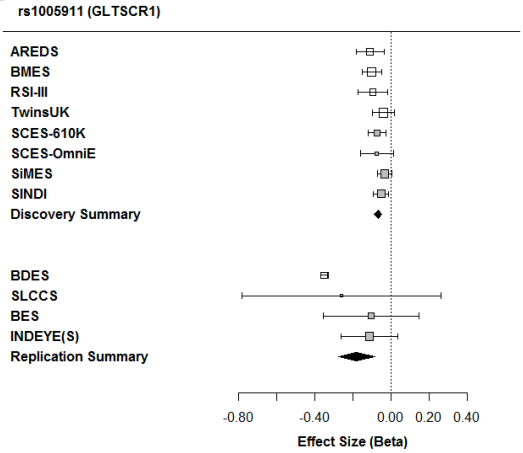
E

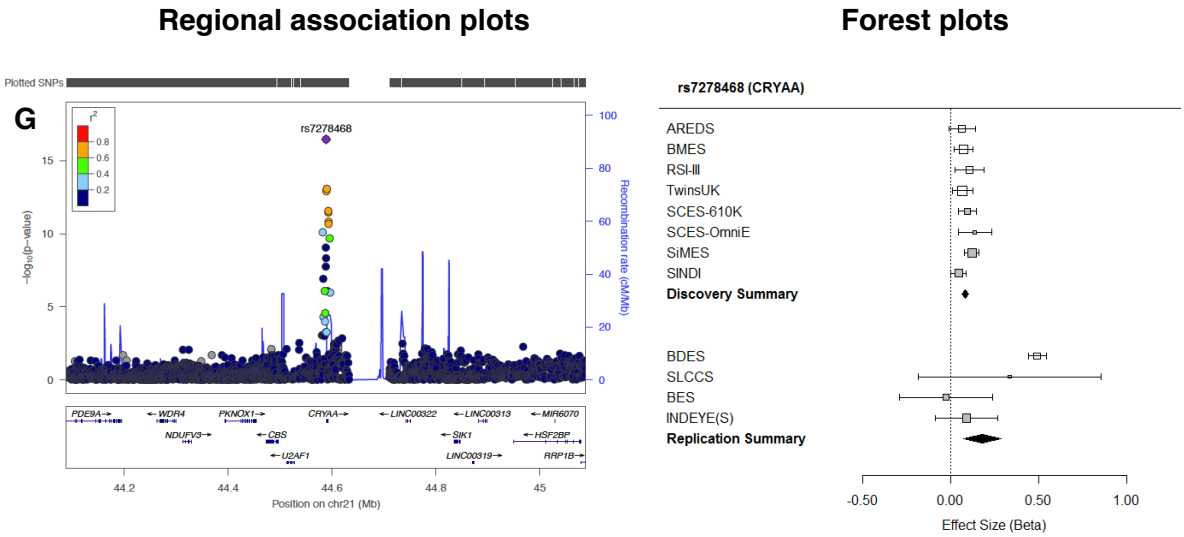




F







This figure shows the regional association and forest plots for the 7 loci association with age-related nuclear cataract at the All Data meta-analysis phase. The regional association plots are centred on the sentinel meta-analysis SNP (purple diamond) per locus. The shading of the adjacent SNPs (±400kb) reflects their pairwise correlation (r^2) with the sentinel SNP. Blue lines — estimated recombination rates (1000G phase3 reference). The forest plots show: the effect size (the linear regression beta from the discovery phase or the ln(odds ratio) from the replication phase) and its standard error (bars). The squares are proportional to weights used in meta-analysis; European ancestry — no shading; Asian ancestry — grey shading. Black diamond — overall effect. A) *COMMD1* locus; B) *LINC01412* locus; C) *SOX2-OT* locus; D) *TMPRSS5* locus; E) *MMAB* locus;F) *GLTSCR1* locus; G) *CRYAA* locus

5.4.4. Common variants in congenital cataract genes

Given that mutations in *CRYAA*, *CRYAB*, *GJA3* and *SOX2* are known to cause congenital cataracts, I explored whether common variants within other congenital cataract genes were associated with age-related nuclear cataract **(Figure 5.8.).** Using publicly available data, I selected 96 genes (α=5x10^{-4}) and found multiple variants at 47 of these to be associated (P<0.05) with nuclear cataract. The most strongly associated variants were in *BFSP1* (P=3.5x10^{-5}), *LIM2* (P=1.4x10^{-4}), *MIP* (P=3.4x10^{-4}*), TFAP2A* (P=3.7x10^{-4}) and *CHMP4B* (P=3.8x10^{-4}). The associated variants within or in proximity to congenital cataract genes explained 2% of variance in age-related nuclear cataract.

Figure 5.8: Common variants in congenital cataract genes.

This Manhattan plot shows the association results for the congenital cataract genes. The -log10(p-value) of the sentinel variant per gene is plotted against the gene location. The colour code represents the strength of association in terms of p-value.

5.4.5. Gene-set enrichment analysis

No pathways or gene sets were statistically significantly enriched after correction for multiple testing. Below are presented the pathway and gene sets that (**Table 5.9**): have connection to cataract formation; were detected by multiple databases; or that were nominally associated at $P \leq 0.01$ according to at least one database.

5.4.6. Functional fine mapping at the associated loci

In order to identify plausible candidate genes within the associated loci, functional fine mapping at each locus focusing on regulation and expression was performed.

Regulatory elements: First, I explored whether the sentinel variant at each associated locus ($P < 1x10^{-5}$), or variants in high LD with it mapped to H3K27ac sites (**Table 5.10**). The presence of H3K27ac accurately marks active enhancers as opposed to inactive/poised enhancer elements containing H3K4me1 modification

alone[575]. The presence of H3K27ac and other enhancer markers (H3K4me1, H3K4me3, H3K9ac) was also queried through HaploReg and if the variant mapped to a binding sequence for any of these, the affected tissues or the number of affected tissues tissues was noted (**Table 5.10**). H3K4me3 and H3K9ac co-localize on active

Table 5.9. Gene set enrichment analysis

		p-value				p-value	
Gene-set	**Databases**	**raw**	**adj.**	**Gene-set**	**Databases**	**raw**	**adj.**
lens structural constituents	GOTERM	3.0E-03	0.86	glucose regulation of insulin secretion	REACTOM	0.01	0.91
ephrin receptor signalling	Ingenuity	0.03	1.00	Glycosphingolipid biosynthesis	KEGG	0.01	0.33
peptidase activity	GOTERM	0.01	0.88	histone deubiquitination	GOTERM	0.01	0.62
proteasome complex	GOTERM	0.01	0.66	Immunity and defence	Panther	0.01	0.55
Proteolysis	Panther	0.01	0.44	integration of energy metabolism	REACTOM	0.01	1.00
Aspartic proteases	GOTERM	0.01	0.72	Metabotropic glutamate receptor group III	Panther	0.01	0.10
	Panther	0.03	0.68	negative regulation of cell growth	GOTERM	0.01	0.76
Cysteine protease	Panther	0.01	0.74	neural crest cell migration	GOTERM	0.01	0.60
ATP-binding cassette (ABC) transporter	GOTERM	0.04	0.98	Neurotransmitter release	Panther	4.0E-03	0.31
	Panther	0.04	0.91	nuclear chromosome, telomeric region	GOTERM	0.01	0.61
cell cycle	BIOCARTA	0.02	1.00	P53 signaling	KEGG	0.02	0.50
	GOTERM	0.01	0.97		REACTOM	0.04	0.70
cell division	GOTERM	3.0E-03	0.68		Panther	0.01	0.17
mitosis	BIOCARTA	0.04	1.00	protein/AA glycosylation	GOTERM	0.01	0.92
	REACTOM	0.01	0.71	protein/AA autophosphorylation	GOTERM	3.0E-04	0.67
	Panther	4.0E-03	0.45	regulation of insuline secretion	REACTOM	0.01	0.90
G1 phase	REACTOM	0.01	1.00	response to hydrogen peroxide	GOTERM	0.01	0.79
dephosphorylation	GOTERM	1.0E-03	1.00	response to starvation	GOTERM	0.01	0.74
electron transport	GOTERM	3.0E-03	0.83	thyroid cancer	KEGG	0.01	0.48
	REACTOM	4.0E-03	0.69	transmembrane receptor regulatory/adaptor protein	Panther	0.01	0.73
	Panther	0.01	0.66	transport	GOTERM	0.003	0.76
endosome	GOTERM	3.0E-03	0.65	ubiquitin ligase complex	GOTERM	0.001	0.64
Extracellular transport/import	Panther	0.01	0.33	WNT signalling	GOTERM	0.01	0.79
G-protein	GOTERM	0.01	0.39		KEGG	0.02	0.65
	Panther	0.01	0.39				

The table shows the results from the gene set enrichment analysis performed using MAGENTA. The gene-set and the database from which the gene-set was obtained as well as the raw and

FDR of 15% adjusted p-values (adj.) for each gene-set are listed.

gene promoters and are associated with active transcription[576]. Similarly to the enhancers it was also noted whether the variant mapped to a DNAse hypersensitivity site at one or more tissues (**Table 5.10**). DNAse hypersensitivity sites are regions of the genome where chromatin has lost its condensed structure, exposing the DNA and making it accessible for transcription factors. Finally, I explored whether presence of the non-reference allele at each sentinel SNP changes any transcription factor binding motifs (**Table 5.10**).

In five cases, sentinel variants fell within clusters of H3K27ac and 50% of sentinel variants mapped to other enhancer marks in many different tissues. Moreover, partitioning the heritability estimated by the LD-score regression resulted in significant enrichment (after adjusting for the number of independent tests) of repressor sequences (0.002) and H3K9ac enhancer picks (0.005). Variants also mapped to a couple of regulatory motifs that may function in lens, cataract or eye development. For example, early embryonic disruption of Klf4 resulted in a smaller lens and nuclear opacity[577]. OTX2 is an important transcription factor for head structure development and mutations in OTX2 cause anophthalmia/ microphthalmia[102,578]. Interestingly, CTCF binding sites were also predicted to be affected. CTCF is one of the key players in the organisation of insulator elements[579]. Finally, according to TiGER, the motifs HaploReg predicted to be affected by the risk allele at the sentinel SNP, point to members of eye specific cis co-regulation networks (POU3F2-FOXJ2-PAX2-PAX6-MEF2-HNF1; SRF-CHX10-PITX2) or cis co-regulation networks with function in the eye (ERalpha-a-GATA6-SMAD3-TGIF, Lmo2_complex-alphaCP1-MEF2-NFY-STATA4). These results may be suggestive of a role for impaired gene regulation in age-related nuclear cataract formation.

Lens expression: Next, I explored whether any of the genes at the associated ($P<1x10^{-5}$) loci were expressed in human and mouse lens, using publicly available data (**Table 5.11**). Multiple transcripts of genes within those loci were expressed in the human eye and or in human or murine lens, but in terms of relative expression the levels were in general low, as measured by both normalised probe intensities (PLIER values) and/or normalised EST counts. However, some of the genes were expressed at moderate to high levels (PLIER/EST* enrichment scores): 12800.8/18.49 for *CRYAA*, for 2860.40/5,46 *CRYAB*, 289.7/1.92 for *SOX2*, 192.9/1.72 for *COMMD1*, 160.5/- for *ZEB2,* for 222.93/0.86 *MAF*. Some of those genes also showed eye-specific

* higher values mean higher expression relative to the mean expression level for genes in the tissues

enrichment (TiGER, **Table 5.11**) and/or were expressed in the human embryonic lens (iSyte, **Table 5.11**): *CRYAA*, *CRYAB*, *ZEB2*, *MAF*, *SOX2*, *GJA3*, *CRX*. Thus, in terms of expression, the above listed genes are plausible functional candidates.

eQTL summary data: Further, the relationship between the alleles present at most sentinel SNP per locus (and the variants in high LD with those (r^2>0.8) and all transcripts within 1MB of the lead SNP were explored (**Table 5.12**). As expression levels in the ocular lens are not available to conduct eQTL analysis, eQTL effects at the associated loci were queried in all of the tissues available from the GTEx database and the TwinsUK Expression data, under the assumption that a statistically significant effect in any of them would demonstrate the capacity of that locus to regulate transcription albeit in other cell lines or tissues. Although effects over transcription may be tissue and age-specific, others are shared across tissues (mean π_1=0.73 of cis-eQTLs)[580,581]. Eight of the SNPs altered expression (or were in LD with the top eQTL) in at least one tissue. For example, the most strongly associated variant in *CRYAA* was also the most significant eQTL for the gene in liver ($P_{liver}=4.8x10^{-8}$, r_{GTEx}=-0.57). Apart from the lens, *CRYAA* is expressed only in liver and kidney but, given the small sample size for kidney in GTEx, reliable eTQL for the gene in kidney could not be established. The sentinel SNP (rs10502150, $P_{fat}=7.6x10^{-22}$, $r_{TwinsUK}$=-0.35) at the *CRYAB* locus in the European cohorts was in perfect LD (r^2=0.95/D'=1.0) with the top eQTL variant (rs11214027) for *CRYAB* in fat. The eQTL results suggest that these polymorphic loci are capable of affecting genetic expression, at least in some tissues and cell lines, which may suggest that the association of these variants with cataract may be due to an eQTL effect.

Table 5.10: Regulatory elements data (ENCODE, HaploReg) for the sentient SNPs associated with age-related nuclear cataract (p-value<1x10^{-5})

Position									
CHR	BP	SNP	Gene	p-value	H3K27ac mark (N)	Enhancer histone marks	DNAse HSS	Binding proteins	Regulatory motifs
1q31.1	190717899	rs147580522	*LOC440704* [b]	4.4E-08	No (0)	-	-	-	-
2p24.1	20747778	rs61185326	intergenic	3.0E-07	No (6)	-	-	-	Foxj2, Foxp1, Mef2, Pou2f2, TATA
2p23.3	24439276	rs13021828	*ITSN2* [b]	6.1E-07	No (0)	-	28 tissues	CTCF, POL2, POL24H8, SRF, STAT3	ERalpha-a, ZBRK1
2p15	62191878	rs62149908	*COMMD1* [b]	6.5E-08	No (0)	9 tissues	BLD, VAS	-	ERalpha-a, Zfx
2q14.1	115778290	rs77791858	*DPP10* [b]	3.1E-06	No (0)	ESDR	KID,BLD	-	HNF1, Irx
2q22.3	145341259	rs16823886	*LINC01412* [d]	8.8E-08	Yes (2)	5 tissues	-	-	-
2q34	211624516	rs36191905	intergenic	6.2E-06	No (0)	SKIN, HRT	-	-	-
2q37.1	235342061	rs10199789	intergenic	7.5E-06	No (17)	-	-	-	Barx1, EWSR1-FLI1, Gbx1, Myf_1, Nkx1-1
3p24.1	30813873	rs1827496	*GADL1* [b]	5.2E-06	No (0)	7 tissues	11 tissues	EGR1	Evi1, Hsf, Ik2, Pax4, Pou2f2, SRF
3q24	148672923	rs76405079	intergenic	9.1E-06	No (4)	GI	-	-	Foxa,NF-kappaB,STAT
3q26.33	181346937	rs9842371	*SOX2-OT* [b]	2.6E-12	Yes (4)	12 tissues	4 tissues	GATA2	Dmbx1,Hoxa5,Otx2
4q22.1	91932333	rs72667512	*CCSER1* [b]	2.6E-06	No (0)	-	-	-	DEC,Lmo2-complex
4q28.1	127728164	rs77573620	intergenic	6.5E-06	No (4)	6 tissues	-	-	Arid5a, Dbx1, HMG-IY, HNF1, Mef2, Ncx, PLZF, Pou2f2, TATA
5q11.22	55870395	rs71624138	*C5orf67* [b]	5.5E-06	No (2)	-	-	-	Pou2f2
5q31.3	139543125	rs156084	*LOC101929719* [b]	9.1E-06	No (6)	BRN	-	-	Ik-1, Ik-2, NRSF, STAT, Sin3Ak-20, Zfx
5q31.3	142111021	rs72796639	intergenic	1.0E-05	No (8)	-	-	-	NF-kappaB,STAT,Spz1
6q25.3	159308438	rs34891821	*C6orf99* [b]	5.1E-06	No (0)	ESDR	-	-	-
6q27	165953127	rs545099	*PDE10A* [b]	9.3E-06	No (6)	-	-	-	BCL, BDP1, UF1H3BETA. Klf4, NF-E2, Pax4, CACD, ZNF219
10q22.1	73199588	rs41281302	*CDH23* [a]	6.9E-06	No (0)	BLD, HRT	-	-	CDP
11q23.1	111788464	rs10789852	HSPB2 [b]/C11orf52 (*CRYAB*)	8.4E-06	No (10)	7 tissues	-	-	BDP1,LUN-1
11q23.2	113566207	rs4936279	*TMPRSS5* [b]	4.2E-11	No (4)	-	-	-	CTCF,E2F, ELF1, HEY1, Rad21, SMC3, Sin3Ak, YY1
11q24.2	126058369	rs487450	intgenic	2.4E-06	Yes (2)	-	-	-	AC-binding-protein, SP1, STAT, TATA, YY1
11q24.3	130270855	rs59317921	*ADAMTS8* [d]	3.4E-06	Yes (5)	-	-	-	Egr1, HMG-IY, NF-AT, Pax5

12q24.11	109988214	rs11067211	*MMAB* [d]	1.6E-07	No (5)	-	IPSC, BLD	-	Pax5,Pbx3
13q12.11	20721897	rs17077135	*GJA3* [b]	6.2E-06	No (2)	-	BRN	-	GR, Sox
14q24.2	73277092	rs1990441	*DPF3* [b]	5.4E-06	No (2)	-	-	-	Irf, PU.1, STAT, TATA
15q23	69096596	rs11639228	*ANP32A* [b]	7.2E-06	No (11)	-	4 tissues	-	Glis2, Myb, RXRA, SP1, SREBP
15q24.2	75449755	rs143061779	intergenic	8.6E-06	No (5)	9 tissues	5 tissues	-	E2A,TBX5,ZEB1
16p13.11	15807137	rs12324990	*NDE* [b] / *MIH11*	2.4E-06	No (2)	7 tissues	4 tissues	NFKB	Hic1,p300
16q23.2	79468817	rs28703582	intergenic	3.1E-06	No (1)	15 tissues	24 tissues	-	-
17p13.3	1654089	rs56679346	*SERPINF2* [b]	8.8E-06	No (3)	GI, MUS, CRVX	CRVX,MUS	-	BDP1, CTCF, Erg1, HES1, Myc, Nanog, SP1, Sin3Ak-20, THAP1
17q25.3	76170735	rs1065769	*TK1* [a]	1.2E-06	Yes (20)	GI	-	-	ERalpha-a,Irf,Pax5
18q21.31	55659197	rs1942539	intergenic	6.5E-06	No (1)	IPSC	-	-	Pou2f2
19q13.33	48206092	rs1005911	*GLTSCR1* [a]	2.8E-07	No (5)	-	-	-	-
21q22.3	44588757	rs7278468	*CRYAA* [c]	3.6E-17	No (8)	IPSC, LIV, PLCNT	-	-	CACD, HDAC2, Klf4, Spz1

This table shows the regulatory elements data (ENCODE, HaploReg) fro the sentinel variants in the combined meta-analysis (Discovery phase). The sentinel SNPs (dbSNP147 rs-number) at each locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; P value – P-value from the Discovery phase meta-analysis (European and Asian cohorts combined); H3K27ac mark – whether a variant falls within a H3K27ac mark (a mark of active regulatory elements); N – H3K27ac marks – number of H3K27ac mark within a 50kb window around the studied variant; Next, four columns summarise whether the sentinel variant maps to other enhancers, DNAse hypersensitivity sites (DNAseHSS), transaction factor binding sites/motifs. If the number of tissues where the regulatory elements are active was under 3 the tissues are listed, else the number of affected tissues is listed. BLD – blood; VAS –vascular; ESDR – embryonic stem cell derivative; KID – kidney; SKIN – skin; HRT – heart; GI – gastrointestinal; BRN – brain; BLD – bladder; MUS – muscle; CRVX – cervix; IPSC – induced pluripotent stem cells, LIV – liver, PLCNT – placenta

Table 5.11: Publicly available expression data in human and mice for the genes closest to or in the same locus as the sentinel SNP

Position								
CHR	BP	SNP	Gene	p-value	iSyTE: lens enriched genes	Ocular tissue database (lens data): PLIER$ scores for genes in the locus	TiGER eye ESTs*: enrichment score, (-log10 p-value)	MGI XGD
1q31.1	190717899	rs147580522	*LOC440704* [b)]	4.4E-08	-	*FAM5C*: 5.50	-	No
2p24.1	20747778	rs61185326	intergenic	3.0E-07	*MATN3*	*MATN3*: 2.2; *WDR35*: 39.31; *HS1BP3*: 31.70; *C2ORF43:* 16.91; *PUM2:* 235.25; *SDC1:* 31.81; *RHOB*: 131.52; *LAPTM4A:* 548.21	MATN3: 0.52, 0.07; *WDR35*: 2.02,1.35; *HS1BP3*: 1.02, 0.26; *C2ORF43:* 1.68, 1.06; *PUM2:* 1.37, 1.04; *SDC1:* 0.79 ,*RHOB*: 0.62, 0.001; *LAPTM4A:* 0.86,0.06	No
2p23.3	24439276	rs13021828	*ITSN2* [b)]	6.1E-07	-	*ITSN2*: 107.6; *KLHL29*: 23.7; *ATAD2B*: 6.9; *UBXN2A*: 170.7; *PFN4*: 26.4; *MFSD2B*: 130.0; *TP53I3*: 183:3; *FKBP1B*: 130.0; *SF3B14*: 138.5	*ITSN2*: 0.61, 0.03; *ATAD2B*: 1.33, 0.53; *PFN4*: 2.88, 0.82; *TP53I3*: 1.32, 0.55; *FKBP1B*: 1.94, 0.94; *SF3B14*: 0.84, 0.15	No
2p15	62191878	rs62149908	*COMMD1* [b)]	6.5E-08	-	*COMMD1*: 192.9, *CCT4*: 292.4; *B3GNT2*: 56.7	*COMMD1*: 1.72, 1.25; *CCT4*: 1.12, 0.63; *B3GNT2*: 0.45, 0.028	No
2q14.1	115778290	rs77791858	*DPP10* [b)]	3.1E-06	-	*DPP10*: 11.7	*DPP10*: 1.22, 0.41	No
2q22.3	145341259	rs16823886	*LINC01412* [d)]	8.8E-08	*ZEB2*	*ZEB2:* 160.5	-	No
2q34	211624516	rs36191905	intergenic	6.2E-06	-	*CPS1*: 15.95; *MYL1*: 34.34; *LANCL1*: 49.64	*MYL1*: 0.56, 0.005; *LANCL1*: 0.66, 0.04	No
2q37.1	235342061	rs10199789	intergenic	7.5E-06	-	-	-	No
3p24.1	30813873	rs1827496	*GADL1* [b)]	5.2E-06	-	*GADL1* 10.4; *TGFBR2*: 21.5	*TGFBR2*: 0.29, 6.0E-05	No
3q24	148672923	rs76405079	intergenic	9.1E-06	*SOX2*	*SOX2-OT*: *289.7; SOX2:* 289.7; *FXR1*: 700.1; *DNAJC19*: 118.4	*SOX2*: 1.92, 1.83; *SOX2-OT*: 2.42, 1.93; *FXR1*: 0.87, 0.12; *DNAJC19*: 0.40, 0.004	No
3q26.33	181346937	rs9842371	*SOX2-OT* [b)]	2.6E-12	*CP, HPS3*	*CP*: 518.2*; GYG1*: 134.6*; HLTF:* 73.2; *HPS3:* 111.3;	*CP*: 2.49, 5.49;*GYG1*; 0.52, 0.006; *HLTF*: 0.42, 0.01; *HPS3*: 0.76, 0.08	No
4q22.1	91932333	rs72667512	*CCSER1* [b)]	2.6E-06	-	-	-	No
4q28.1	127728164	rs77573620	intergenic	6.5E-06	-	-	-	No
5q11.22	55870395	rs71624138	*C5orf67* [b)]	5.5E-06	*IL6ST*	*ANKRD55*: 167.80; *IL6ST*: 285.54; *DDX4*: 8.48; *IL31RA*: 12.84;	*IL6ST*: 0.81, 0.06; *DDX4*: 0.47, 0.05	No
5q31.3	139543125	rs156084	*LOC101929719* [b)]	9.1E-06	*HBEGF, PFDN1, CXXC5, DNAJC18*	*HBEGF*: 65.5 *PFDN1*: 96.0 *CXXC5*: 65.5;*DNAJC1*: 26.1	*HBEGF*: 0.38, 0.01; *PFDN1*: 0.27, 1.8E-05; *CXXC5*: 1.31, 0.98; *DNAJC1*: 0.80, 0.11	No
5q31.3	142111021	rs72796639	intergenic	1.0E-05	-	*FGF1*:32.40, *ARHGAP26:* 49.33; NR3C1: 88.27	*FGF1*:0.72, 0.09, *ARHGAP26:* 0.73, 0.040 NR3C1: 0.57, 0.009	No
6q25.3	159308438	rs34891821	*C6orf99* [b)]	5.1E-06	*DYNLT1*	*DYNLT1*: 415.01; *RSPH3*:48.11; *TAGAP*: 17.31; *EZR*: 36.32;	*DYNLT1*: 0.29, 4.3E-05; *TAGAP:* 1.12, 0.27; *FNDC*: 1.78, 0.63; *SYTL3*: 0.47, 0.05; *TMEM181*: 0.72, 0.05	No
6q27	165953127	rs545099	*PDE10A* [b)]	9.3E-06	-	*PDE10A:* 26.4	PDE10A: 0.39, 0.034	No
10q22.1	73199588	rs41281302	*CDH23* [a)]	6.9E-06	*UNC5B*	*CDH23*: 31.2;*UNC5B*: 30.4, *SLC29A3*: 38.6	*CDH23:* 3.67, 8.59; *UNC5B:* 1.21, 0.44; *SLC29A3*: 2.93, 2.48	No
11q23.1	111788464	rs10789852	HSPB2 [b)]/C11orf52 (*CRYAB*)	8.4E-06	*CRYAB, HSPB2, C11orf52, ALG9*	*CRYAB:* 2860.4*; HSPB2:* 273.6; *C11orf5:* 275.6; *ALG9:* 30.1; *FDXACB1*: 30.1; *DIXDC1*: 188.6; *PIH1D2*: 9.0; *DLAT*: 68.3; *TIMM8B*: 431.7; *SDHD*: 9.0; *PPP2R1B*: 63.0	***CRYAB*: 5.46, 93.22**; *HSPB2*: 1.22; 0.31; *ALG9*: 0.87, 0.16; *DIXDC1*: 1.02, 0.26; *DLAT*: 1.15, 0.43; *TIMM8B*: 1.13, 0.36; *SDHD*: 0.40, 7.0E-07	*Cryab*
11q23.2	113566207	rs4936279	*TMPRSS5* [b)]	4.2E-11	*TMPRSS5*	*TMPRSS5:* 14.1*; ZW10:* 32.0; *USP28*: 38.9; *HTR3B*: 14.6; *HTR3A*: 24.2; *ZBTB16*: 41.8	*ZW10:* 0.25, 0.008; *USP28*: 0.30; 0.02; *HTR3A*: 1.50, 0.41; *ZBTB16:* 1.04, 0.30	No

11q24.2	126058369	rs487450	intgenic	2.4E-06	*SRPR, TIRAP*	*SRPR:* 51.2; *TIRAP:* 30.0; *CDON*: 22.1*; RPUSD4*: 51,7; *DCPS*: 42.6; *ST3GAL4*: 60.0; *FOXRED1*: 25.8; *KIRREL3*: 60.0	*SRPR*: 0.54, 0.001; *TIRAP*: 0.54, 0.03; *CDON:* 1.50, 0.56; *RPUSD4*: 0.88, 0.15; *DCPS:* 0.69, 0.08; *ST3GAL4*: 0.55, 0.02; *KIRREL3*: 1.32; 0.67	No
11q24.3	130270855	rs59317921	*ADAMTS8* [d]	3.4E-06	-	*ADAMTS8*: 31.5; *ADAMTS15*: 21.5	*ADAMTS8*: 0.72,0.05; *ADAMTS15:* 0.99, 0.22	No
12q24.11	109988214	rs11067211	*MMAB* [d]	1.6E-07	-	*MMAB*: 76.1; *FOXN4*: 23.0; *MYO1H*: 19.6; *KCTD10*: 50.3; *UBE3B*: 78.9; *MVK*: 31.5	*MMAB*: 0.79,0.13; *KCTD10*: 1.35, 0.78; *UBE3B:* 1.48, 1.27; *MVK*: 1.44, 0.84	No
13q12.11	20721897	rs17077135	*GJA3* [b]	6.2E-06	*GJA3*	*GJA3*: 74.6	-	*Gja3*
14q24.2	73277092	rs1990441	*DPF3* [b]	5.4E-06	*DFP3*	*DPF3*: 42.7;*DCAF4*: 29.1; *RBM25*: 58.6; *ZFYVE1*: 40.5; *PSEN1*: 82.1; *PAPLN*: 42.8; *NUMB*: 143.3; *ACOT2*: 19.0; *ACOT4*: 17.8; *HEATR4*: 28.1; *PNMA1*: 206.0; *DNAL1*: 19.4;	*RBM25*: 0.601, 0.012; *ZFYVE1*: 1.89, 1.40; *PSEN1*: 0.60, 0.01; *PAPLN*: 1.36, 0.56; *NUMB*: 0.35, 4.6E-04; *ACOT2*: 0.16, 1.2E-05; *PNMA1*: 1.87, 2.50; *DNAL1*:1.08, 0.32	No
15q23	69096596	rs11639228	*ANP32A* [b]	7.2E-06	*ANP32A*	*ANP32A*: 57.6; *SPESP1*: 28.6; *NOX5*: 28.5	*ANP32A*: 0.99, 0.26;	*Anp32a*
15q24.2	75449755	rs143061779	intergenic	8.6E-06	*SCAMP5*	*SCAMP5*: 35.2; *PPCDC*: 22.5;	*SCAMP5*: 2.34, 2.98; *PPCDC*: 0.67, 0.08	No
16p13.11	15807137	rs12324990	*NDE* [b] */MIH11*	2.4E-06	*NDE1*	*NDE1*: 19.6; *MYH11*: 19.6; *MPV17L*: 26.4	*NDE1*: 0.55, 0.02; *MYH11*: 1.08, 0.52;	No
16q23.2	79468817	rs28703582	intergenic	3.1E-06	*MAF*	*MAF*: 222.93	*MAF*: 0.82, 0.12	No
17p13.3	1654089	rs56679346	*SERPINF2* [b]	8.8E-06	*SGSM2, PRPF8, YWHAE*	*SERPINF2:* 25.0*; SGSM2: 32.1 PRPF8:* 160.2; *YWHAE:* 1658.2*; WDR81:* 27.5; *SCARF1*: 55.8; *SLC43L2*: 24.1; *RILP*: 19.8; *PTPNA*: 84.5; *MYO1C*: 39,6 *INPP1K*: 22.4; *CRK*: 115.7; *TUSC5*: 30.0;	*SERPINF2:* 0.89, 0.18*; PRPF8:* 1.41, 2.27; *YWHAE:* 1.13, 0.79; *WDR81:* 0.25, 0.008; *MYO1C*: 0.62, 0.0061; *CRK*: 1.38, 1.05;	*Ywhae*
17q25.3	76170735	rs1065769	*TK1* [a]	1.2E-06	*TMC6, SYNGR2*	TK1: 45.6; *TMC6:* 42.6; *SYNGR2*: 28.1; *TNRC6C*: 71.2; *BIRC5*: 42.8; *TMC8*: 37.8	TK1: 0.88, 0.12; *TMC6:* 0.28, 0.0005; *SYNGR2*: 0.13, 2.68E-07; *TNRC6C*: 2.89, 3.26; *BIRC5*: 1.18, 0.61;	No
18q21.31	55659197	rs1942539	intergenic	6.5E-06	-	*NEDD4L*: 70.69; *FECH*: 34.31; *NARS*: 339.601; *ATP8B1*: 14.60	*NEDD4L*: 1.21, 0.73; *FECH*: 0.93, 0.20 *NARS*: 0.98, 0.24; *ATP8B1*: 1.92, 1.45	No
19q13.33	48206092	rs1005911	*GLTSCR1* [a]	2.8E-07	*CRX*	*GLTSCR1*: 55.6; *KPTN:* 20.7; *SEPW1*: 164.1; *GLTSCR2*: 69.8; *CRX*: 14.3; *NAPA*: 45.0; *TPRX1*: 18.0; *ZNF451*: 92.0; *SNORD23*: 69.8;	*GLTSCR1*: 0.34, 3.3E-04; *KPTN:* 1.68, 0.84; *SEPW1*: 1.05, 0.40; ***GLTSCR2*: 2.04, 13.90**; ***CRX*: 18.21, 74.57**; *NAPA*: 1.23, 0.91; *ZNF451*: 0.78, 0.08;	No
21q22.3	44588757	rs7278468	*CRYAA* [c]	3.6E-17	*CRYAA*	*CRYAA:12800.8; UTAF1:* 119.4	***CRYAA*: 18.49, 329.77**	*Cryaa*

This table summarises publicly available expression data in human and mice for the genes closest to or in the same locus as the age-related nuclear cataract sentinel SNP (p-value <1x10^{-5}, Discovery phase combined meta-analysis. The sentinel SNPs (dbSNP147 rs-number) at each locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; P value – P-value from the Discovery phase meta-analysis (European and Asian cohorts combined); Locus here means a region between two recombination hot-spots (recombination rates ≥ 40) or a 1MB window, whichever was smaller; iSyTE – integrated Systems Tool for Eye gene discovery (http://bioinformatics.udel.edu/Research/iSyTE); Ocular tissue database (https://genome.uiowa.edu/otdb/): $PLIER score – normalised score from the probe logarithmic intensity error algorithm over all exones of a transcript; higher scores represent higher expression of a transcript; * TiGER score – score for enrichment of expressed sequencing tags (ESTs), normalised over the size of the EST library; higher scores represent higher expression of the transcript (scores > 5 denote tissues specific genes; bold in table); Gene that were not present in TiGER or had a score of 0 for enrichment in the eye were omitted. MGI GXD – Mouse Genome Informatics Gene Expression Database (http://www.informatics.jax.org/expression.shtml).

Table 5.12: GTEx and TwinsUK eQTL summary data

Position										LD between sentinel SNPl and eSNP		sSNP – eSNP distance (in kb)
CHR	BP	sSNP	Gene	p-value	N	eGene	eTissue	eSNP	eQTL p-value	LD(r^2/)	D'	
2p24.1	20747778	rs61185326	intergenic	3.0E-07	1	*RHOB*	Muscle (Skeletal)	rs56900434	3.0E-07	0.89	1.0	1.5
2p23.3	24439276	rs13021828	*ITSN2* [b)]	6.1E-07	11	*AC009228.1*	Esophagus - Muscularis	rs10195436	1.2E-06	0.87	0.96	3.4
						AC008073.7	Testis	rs76172309	5.1E-08	0.96	1.0	19.9
11q23.1	111788464	rs10789852	HSPB2 [b)]/C11orf52 (*CRYAB*)	8.4E-06	22	*DLAT*	Muscle (Skeletal)	rs10789852	1.5E-09	-	-	0.0
						C11orf52	Fat	rs4252591	3.7E-16	1.00	1.0	18,485
						*CRYAB***	Fat	rs11214027	7.6E-22	0.76	0.9	18,485
						PIH1D2	LCL	rs11214075	1.4E-06	0.83	1.0	18,485
						PPIHP1	Liver	rs11214075	2.5E-06	0.83	1.0	18,485
11q24.3	130270855	rs59317921	*ADAMTS8* [d)]	3.4E-06	1	*RP11-121M22.1*	Testis	rs11222084	3.8E-08	0.93	1.0	2.6
12q24.11	109988214	rs11067211	*MMAB* [d)]	1.6E-07	5	*MMAB*	Brain Cerebellum	rs11836136	5.3E-08	0.96	1.0	18.5
							Brain - Cortex	rs754179	1.3E-06	0.96	1.0	16,1
15q24.2	75449755	rs143061779	intergenic	8.6E-06	6	*CTD-3154N5.1*	Nerve - Tibial	rs143061779	6.9E-06	-	-	0.0
17p13.3	1654089	rs56679346	*SERPINF2* [b)]	8.8E-06	8	*SERPINF2*	Pancreas	rs12051752	1.5E-18	0.85	1.0	16.5
17q25.3	76170735	rs1065769	*TK1* [a)]	1.2E-06	1	*SYNGR2*	Thyroid	Yes	2.2E-06	-	-	0.0
						RP11-712L6.7	Skin	rs598169	1.6E-06	0.91	1.0	chr11
							LCL	rs1695739	8.5E-12	0.81	1.0	chr11
						SRPR	LCL	rs8177376	1.6E-10	0.91	1.0	chr11
21q22.3	44588757	rs7278468	*CRYAA* [c)]	3.6E-17	2	*CRYAA*	Liver	rs7278468	4.8E-08	-	-	0.0

The table reports the summary eQTL data obtained from GTEx and TwinsUK (purple) for the sentinel SNP at each locus.For GTEX only cis effects are shown. For TwinsUK both cis and trans-effect are reported. The sentinel SNPs (dbSNP147 rs-number) at each locus is listed as follows: position — chromosomal band (CHR) and base pair position (GRCh37/hg19, positive strand); Gene — closest RefSeq gene (± 2kb from the longest transcript start/end sites); variant location relative to gene: a) exonic/UTR; b) intronic; c) upstream; d) downstream; P value – P-value from the Discovery phase meta-analysis (European and Asian cohorts combined); Locus here means a region between two recombination hot-spots (recombination rates ≥ 40) or a 1MB window, whichever was smaller; SNPs with no eQTL effect and SNPs not in LD (r^2>0.8) with the top eQTL per gene are omitted. eGene – the gene whose expression is dependent on sentinel SNP, eSNP – the eQTL SNP for each gene. eTissues – the tissues where the eSNP exerts its effect on expression; N –number of eQTL-eGene-eTissues associations; LD between sentinel SNP and eSNP is reported as calculated on bases of 1000 Genomes LD parameters (r^2 and D' according to SNAP (https://www.broadinstitute.org/mpg/snap/ldsearchpw.php) or LD calculator (https://caprica.genetics.kcl.ac.uk/~ilori/ld_calculator.php; in the cases SNP were > 500kb apart)), as is the distance between the two SNPs in kilo-bases (kb); LCL: lymphoblastoid cell line;

5.4.6. Candidate genes identification

At each locus (p-value < $1x10^{-5}$), whenever possible, a candidate gene or a set of candidate genes was identified. This was done on bases of:

- the bioinformatics work described above;
- functional work in samples from human lens nuclei (RNA sequencing and protein expression and localisation work) that was carried by our collaborators in the labs of Dr. Chaolong Wang and Dr. Anita Chan (data not shown);
- extensive search of the published literature to date.

The reasoning behind selecting candidate genes at each locus will be described here. Loci that did not replicate or where there was insufficient data to pinpoint a likely candidate are omitted.

***COMMD1* locus**: There were several genes mapping to the *COMMD1* locus, a locus of extended high-LD (**Figure 5.7.**). The publicly available expression data and the eQTL summary data were not informative for this locus but the sentinel SNP was predicted to disrupt an estrogen receptor binding site with function in the eye (**Table 5.10.**). Two of the genes in the locus (*COMMD1* and *EHBP1*) were associated with cataract severity in the RNA sequencing data provided by of Dr. Wang and the immunocytochemistry analysis provided by Dr. Chan suggested reduction in

expression levels of both genes. The CORD analysis showed *COMMD1* to be co-expressed with many of the subunits of the mitochondrial protein synthesis machinery, while *EHBP1* was co-expressed with cell adhesion proteins. Both genes were treated as plausible candidates at this locus.

***LINC01412* locus:** The sentinel variant at this locus mapped to an enhancer active at 5 different tissues (**Table 5.11**). From the genes in the *LINC01412* locus, *ZEB2* had the highest expression(**Table 5.10**) in human lens and its expression was elevated with increased cataract severity according to the RNA sequencing analysis. CORD analysis showed *ZEB2* to be co-expressed with extracellular matrix receptor and focal adhesion proteins (including VIM: p-value=6.8×10^{-36}). *VIM* encodes for a type III intermediate filament protein and mutations in this gene are a known cause of Mendelian cataracts.

***SOX2-OT* locus:** The sentinel variant at the *SOX2-OT* was located within intron 5 of *SOX2-OT* gene, a non-coding RNA gene that regulates the expression of *SOX2*[582]. As described above, the sentinel variants in not predicted to be an eQTL (**Table 5.12.**) but it is predicted to fall within a site for OTX2 transcription factor binding site. *SOX2* was expressed in the lens (**Table 5.10.**) but, according to the RNA sequencing results, its expression in lens epithelium was not associated with cataract severity. However, SOX2 protein concentration was elevated in epithelial nuclei of cataractous lenses compared to control lenses.

***CRYAB* locus:** There were again many genes in high-LD at this locus (**Figure 5.7**). However, the only one strongly expressed in lens was *CRYAB* (**Table 5.10.**). The sentinel variant at this locus was the eQTL SNP for *CRYAB* in fat (**Table 5.12.**). *CRYAB* encodes for a subunit of the most abundant structural lens protein (α-Crystallin).

***GJA3* locus:** This locus showed complicated recombination patterns but only one gene, *GJA3*, was in LD (recombination rates <20). The gene was expressed in lens (**Table 5.10.**) and the sentinel variant was predicted to disrupt the binding of members of the SOX family of transcription factors. *GJA3* encoded for a structural lens protein and as discussed further in this chapter, mutations in this gene cause Mendelian forms of cataract[60,583,584].

***GLTSCR1* locus:** Pinpointing a likely candidate in at the *GLTSCR1* locus was not straightforward. The sentinel variant was not predicted to affect gene regulation (**Table 5.10.**) or to be an eQTL SNP (**Table 5.12.**). According to publicly available data *GLTSCR2* and *CRX* showed eye-specific expression, while *CRX* was expressed in the human embryonic lens (**Table 5.11.**). The RNA sequencing analysis showed that *GLTSCR1* expression was not associated with cataract severity but the expression of

both *GLTSCR2* and *NAPA* decreased with increase in cataract severity. There was also a shift from cytoplasmic to nuclear *NAPA* expression in the cataractuos eye. The CORD analysis did not yield any additional information either. *CRX* encodes for a transcription factor, *NAPA* encodes for a vesicle docking protein and the protein encoded by *GLTSCR2* is involved in ribosomal biogenesis. Additional functional work is needed to more accurately pinpoint the causative gene at this locus.

***CRYAA* locus:** At this locus only one gene, *CRYAA*, was in LD with the sentinel variant (**Figure 5.7.**). This gene showed eye- and lens-specific expression (**Table 5.11.**). As described above, the sentinel variant at this locus was also the eQTL SNP for *CRYAA* in liver (**Table 5.12.**). Furthermore, the sentinel variant at this locus is predicted to fall within Klf4 regulatory motifs (**Table 5.10.**). Klf4 is a transcription factor that important for lens maturation[577]. The RNA sequencing analysis suggested that increase in cataract severity score was associated with decrease in *CRYAA* expression, while the immunohistochemistry assays suggested differences in localisation of CRYAA protein between cataractous and normal lenses. All this, together with the fact that *CRYAA* encodes for a subunit of the most abundant structural lens protein, points to *CRYAA* having a function in the formation of ARC.

5.5. Discussion

In this chapter I have presented the largest to date meta-analysis of GWAS for age-related nuclear cataract. As estimated using GCTA and LD-score regression, the variants identified through the meta-analysis explained 9% of phenotypic variance in age-related nuclear cataract, of which 3% was attributable to the common variants that achieved p-values < $1x10^{-5}$ and 2% to genetic variants within congenital cataract loci. I also explored whether there was genetic sharing between age-related nuclear cataract and any of the traits (N = 219) for which GWAS results were available through LDHub, a platform that also uses LD-score regression. However, no statistically significant genetic sharing between nuclear cataract and LDHub phenotypes was detected, probably due the fact the LDHub uses only a subset of variants for the analysis (only the common HapMap 3 imputed SNPs as opposed to the full 1000G imputed set of SNPs).

After replication, 7 loci were associated with age-related nuclear cataract (**Table 5.8.**, **Figure 5.7.**) at genome wide significance level and will be discussed below. Many more (p-value < $1x10^{-5}$) loci contained genes that are potentially interesting functional candidates and will also be discussed here. In terms of function, the candidate genes

identified in this chapter can be divided broadly into two groups: structural lens proteins and transcription factors regulating the expression of those.

Little is known about the genetic aetiology of ARC but oxidation is thought to play a role in nuclear cataract formation through the denaturation/aggregation of crystallins — the structural lens proteins[126,585]. In this study, variants in proximity to *CRYAA* (p-value = 3.6×10^{-17}) and *CRYAB* (p-value = 8.4×10^{-6}), encoding for the two forms of α-crystallin, were associated with age-related nuclear cataract. α-Crystallins contribute to the clarity and refractive properties of the lens in two ways: when bound they are an integral part of the lens protein structures; in the free state they are molecular chaperones that prevent protein damage and protect against oxidative stress[370,586,587]. Moreover, Mendelian mutations in both genes cause congenital cataracts[57,59,588,589]. *CRYAA* and *CRYAB* are proposed to be down-regulated in lens epithelia of patients with age-related nuclear cataract[27,590,591]. If this is also true during nuclear fibre formation, that would imply that people who have lower expression of crystallins are likely more susceptible to oxidative stress-related protein depletion. The eQTL results we obtained for those two genes together with the lens epithelium expression data are also in agreement with a crystallin depletion hypothesis. Further modulators of crystallin expression, such as SOX2 may also play a role. *SOX2-OT* (p-value = 4.8×10^{-20}) encodes for a highly conserved long noncoding RNA expressed in the lens[582]. *SOX2-OT* regulates a transcription factor (*SOX2*) with an important role for lens and eye development, which when mutated, is known to cause microphthalmia[592-594]. During murine lens placode formation, *Sox2* is regulated by Six3 and Pax6; and combination of Pax6/c-Maf with Sox2 generates synergistic effects in crystallin regulation[66,595]. In chicks Sox2 regulates δ-crystallins expression[596]. Another transcription factor, *ZEB2* (a.k.a. *SIP1*), identified here, suppresses E-cadherin, up-regulates *FOXE3* and is important for fibre cell migration and organisation and cataract formation[79,597,598].

The results also support a role for protein clearance — a mechanism previously suggested to be important for cataract formation[145,599]. *GJA3* (p-value = 6.2×10^{-6}) encodes for a gap-junction connexin (Connexin-46, CXA46) involved in intercellular communication and cellular trafficking; and rare mutations in the gene cause congenital cataracts[60,583,584]. In animal models, alterations in ubiquitination lead to depletion of CXA46, accumulation of CXA43 and increase in Ca^{2+} levels and cataract[600]. COMMD1 (p-value = 5.2×10^{-8}) regulates, among others, copper homeostasis and is involved in ubiquitination of various substrates[601,602]. Copper concentration is higher in ARCs[159]. The *COMMD1* locus contains another plausible candidate gene *EHBP1* the expression

of which showed even stronger association with cataract severity. EHBP1 may play a role in endocytosis of glucose and lipids and is a binding partner of EHD1 — a protein required for lens development[603-605]. The candidate genes at the *COMMD1* locus seem to be under estrogen control, and the RNA sequencing analysis pointed to *COMMD1* expression differences between females and males with opposite direction of effect in females to males. These results need further refinement but if true are in support of the estrogen hypothesis of cataract formation discussed in detail in **Chapter 1**.

The strengths and limitation of the study are discussed below. This study used the largest sample to date for genetic analysis of nuclear cataract and, in the discovery phase, of precisely and quantitatively phenotyped cohorts, resulting in the identification of 6 new loci associated with age-related nuclear cataract and the replication of the association with *CRYAA* variants. Although this is the biggest study to date many more genetic loci will be discovered by further increasing the sample size, as power to detect genetic associations is a function of the number of individuals included in the analysis. We also could not detect statistically significantly enriched gene-sets or pathways, possibly because of the relatively small number of known genetic loci compared to the potentially plurigenic architecture of nuclear cataract. This can also be remedied by better-powered analyses.

Another power-limiting source is heterogeneity. The GWAS used in the study employed a variety of grading systems, which may have increased heterogeneity and therefore, decreased power to detect associations. To minimise the importance of phenotypic heterogeneity to the results the scores were standardised prior to the GWAS. Similarly, the population diversity inherent to a multi-ethnic meta-analysis design[606], presents with both advantages and disadvantages. While replication of association signals among individuals with different ethnic background leads to more precise results and increases the confidence in the veracity of those results, variation in risk allele frequency or underlying differences in linkage disequilibrium between different ancestral groups may decrease statistical power. For the same reason, although the conditional analysis did not suggest the presence of secondary signals in the genome-wide associated loci, this result needs to be taken with caution, as the analysis depended on the assumption that the 1000G LD data are able to capture the LD patterns in the participating cohorts. The 1000G participants are predominately of European dissent and the Asian ancestry populations are not well covered.

Another important contribution of the study is that it provides evidence of sharing of genetic mechanisms between congenital and ARC, and shows the importance of common genetic variants in maintaining crystalline lens integrity in the ageing eye.

Several congenital cataract genes where common variants were associated with age-related nuclear cataract, deserve particular attention in the future. In particular the structural lens proteins: α-crystallins and the β-crystallin and γ-crystallin clusters, where variants were nominally associated with nuclear cataract, as well as *GJA3*, *SOX2*, *BFSP1*, *LIM2*, *MIP, TFAP2A* and *CHMP4B*.

Not all associated loci ($P<1x10^{-5}$) candidate genes known to relate to cataract formation or lens biology (eh. *GLTSCR1* locus) and, therefore, functional work that can shed light on the likely candidate genes in those loci and their role in lens function and cataract formation will be worthwhile.

Another strength of this study is the use of human lens samples for RNA and protein analysis which show a link between genetic variation and transcription regulation, suggesting a role for the latter in age-related nuclear cataract causation. Other aspects of the functional work presented here were limited by the lack of lens tissues. Although, the results showed a likely consistency of findings in all tissues studied, a contextual interpretation of results, taking into account issues of tissue specificity, will be necessary.

Chapter 6. Dietary factors and cataract progression. Diet-related changes in metabolites and gut flora in relationship with cataract.

6.1. Overview

In Chapter 1, I discussed in detail the role of dietary factors in cross-sectional cataract prevalence and in cataract progression and concluded that, we know relatively little about the role dietary factors play in cataract progression. In this chapter, it was also noted that, although the roles of separate nutrients are well studied, diets as a whole are not. Finally, it is also not known whether the effect of dietary nutrients or diets on cataract is direct or mediated by changes in metabolism or in gut microbiome composition. Therefore, in this chapter I aimed to fill these gaps in the current knowledge by:

a) applying a classical twin model to separate the role of genetic and environmental factors in cataract progression, followed by conducting an epidemiology study to evaluate the role of dietary micronutrient intake on cataract progression.

b) conducting an epidemiology study on the relationship between healthy diet, as represented by two different indexes (Mediterranean diet score (MDS), Healthy Eating Index (HEI)), and cross-sectional cataract measurements.

c) assessing the blood metabolome and the gut microbiome associated with habitual food intake; and exploring whether the association between diet and cataract is mediated by the associated metabolites or the gut microbes.

Given the wide variety of dietary scores[607,608] it is necessary to briefly introduce the rationale behind choosing to focus on the MDS and HEI. Next metabolomic and microbiome analysis will be introduced in relation to the health of the eye and the lens. The different hypotheses underlying the analysis presented in this chapter will also be discussed.

6.2. Introduction

The Mediterranean diet has been widely promoted as one of the healthiest dietary patterns, not least because of it being rich in antioxidants. This diet is characterised by high consumption of vegetables and olive oil and moderate consumption of protein, and is based on the traditional foods and drinks of the countries surrounding the Mediterranean Sea[609]. The protective effect of this diet

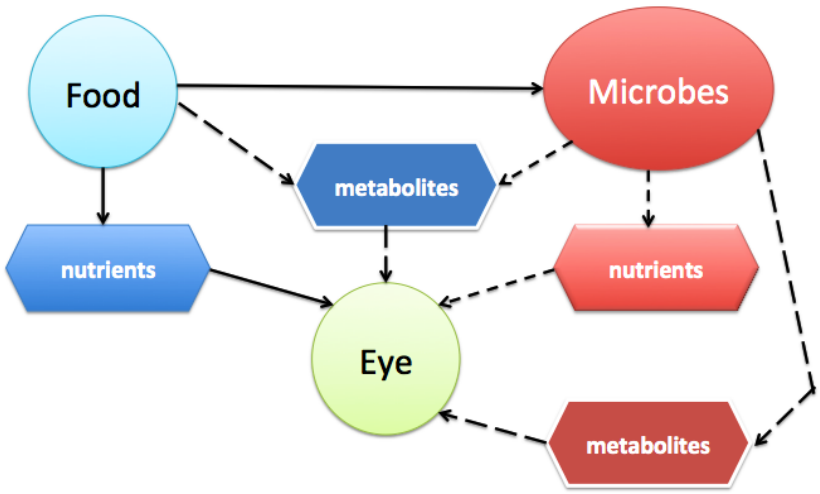
against chronic disease, including eye conditions[610-612] has been extensively studied. The Mediterranean diet has been robustly associated with reduced the risk of overall mortality, cardiovascular diseases, coronary heart disease, myocardial infarction, overall cancer incidence, neurodegenerative diseases and diabetes[609]. Although adherence to Mediterranean-type diet is a plausible candidate in terms of cataract protection, the TwinsUK cohort individuals are mostly elderly British women adhering to a more traditional British Diet. For this reason an additional diet score (the Healthy Eating Index), which is perhaps better suited for studying a western European diet, was evaluated[613]. This score (range 0-100), originally developed in the USA, has ten components of which the first five are nutrient intake per major food group (grains, fruits, vegetables, meat and alternates and milk) and the next five relate to dietary guidelines (total fat, saturated fat, cholesterol, sodium and variety). The score correlates with a range of nutrient intakes, with the person's body mass index (BMI) and with the individual's self-perception of their diet quality[613].

The effect of diet on cataract can be either direct, meaning changes in concentration of certain micronutrients (e.g. vitamins) affect the lens physiology leading to cataract (**Figure 6.1.**), or indirect, for instance nutrients would change the concentration or types of metabolites in blood which in turn could affect the lens. Metabolites are low-weight molecules that are a products of or intermediates of metabolic processes. From a systems biology prospective, metabolites are thought to be closer to the phenotype of interest than either genetic variation or mRNA/protein expression; and may also be affected by environmental influences such as diet, lifestyle choices, disease-related changes, medication, etc[614,615]. Thus metabolites are a powerful tool for identifying diagnosis, prognosis, prediction or treatment efficacy biomarkers[616,617].

Metabolomics is a high-throughput technology that captures the global metabolic state of an individual at a given point in time by simultaneously assaying an extensive set of metabolites[614]. Metabolomics profiles can be obtained from a variety of different biological samples, with blood and urine being the two most often used. There are two types of approaches used for metabolome analysis. The first is an untargeted hypothesis-free approach where hundreds of metabolites are measured, including unknown metabolites, with the idea of obtaining a global view of the metabolome and detecting previously unsuspected or unknown metabolic perturbations associated with a certain disease and novel mechanisms or pathways, akin to GWAS[615]. The disadvantage is the semi-quantitative nature of the methods and the need to validate

any identified compounds. Various methodologies, including clustering, pathway analysis and the use of genetic information can help with the identification of unknown compounds. Targeted approaches rely on small, predefined sets of metabolites and typically focus on a few pathways of interest[615]. The advantages of this kind of approach, apart from easier identification of compounds, include a higher degree of sensitivity (low detection limit) than untargeted approaches and most importantly absolute instead of relative quantification[618].

Figure 6.1. Direct and indirect effects of diet on the eye



The figure shows the possible direct (solid lines) and indirect (dotted line) effects of dietary foods on the eye.

The main methods employed for detecting metabolites in biological samples are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). NMR is a high-throughput, high-speed, low-cost technology that provides information on molecular structures based on atom-centred nuclear interactions[619]. NMR detection is quantitative, non-discriminating, non-destructive, highly reproducible approach for metabolite detection that requires minimal* sample preparation[619]. However, NMR spectroscopy has a relatively low level of sensitivity and can detect only metabolites at medium-to-high levels of abundance[620]. MS-based profiling is also quantitative but it is highly-selective, sensitive and comprehensive with the potential to identify individual

* in the case of soluble molecules

metabolites[621]. Prior to MS, metabolites within a sample usually* need to be separated (e.g. using gas (GC) or liquid (LC) chromatography) which requires long and elaborate preparation that can result in metabolite loss[622].

Little is known about the eye metabolome, its similarities and differences with the blood metabolome or how circulating metabolites would affect it, especially given the presence of aqueous-blood and retina-blood barriers. Studies in animal models have, however, shown the utility of exploring the changes in metabolites in context of eye conditions[623]. Studies in humans are generally lacking but the few that exist presented encouraging results. The metabolite profile of the lens differs from that of other eye tissues, which may be related to the fact that the aqueous humor is the main nutrient source for the lens[624]. The most abundant aqueous humor metabolites, however, seem also to be present in serum at either similar or lower concentration to the aqueous humor[625]. Serum and vitreous humor metabolite profiles, have also been useful in distinguishing case from controls in eye conditions with known systemic risk factors (e.g. neovascular AMD, diabetic retinopathy, glaucoma, etc.)[626-628].

Another possible indirect mechanism is that diet causes changes in the type of bacteria present in the gut and those changes cause systemic effects that extend to lens (**Figure 6.1.**). The gut hosts almost 100 trillion microorganisms (bacteria, archaea and viruses) that share symbiotic properties with humans[629]. It is estimated that these organisms contain in total close to 10 million genes and only about 36% of those are commonly (> 5%) shared between humans[630]. However, there is a substantial redundancy in bacterial genes and functions[631]. At present there are two methods for measuring the metagenome: targeted sequencing of nine highly variable regions of the bacterial 16S ribosomal RNA (rRNA) gene, or shotgun whole genome sequencing[632]. The latter method is more expensive and computationally intensive but it is better at detecting bacterial species and bacterial diversity and is better at bacterial gene prediction[632].

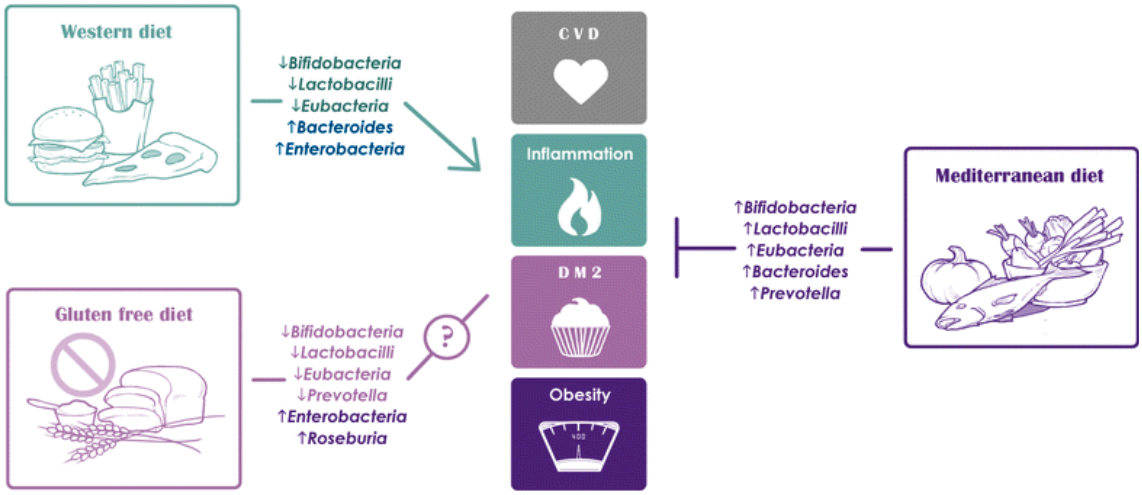
Intestinal microbiota regulate part of the host's metabolic and energy balance, including regulation of cytokines' secretion from adipose tissue and insulin signalling[629]. Furthermore, gut bacteria produce a substantial proportion (90-95%) of the body's short-chain fatty acids, help humans digest plant fibres and bio-activate lignans and flavonoids[633,634]. The gut microbiome also modulates intestinal motility, regulates immune system maturation and confers protection against pathogens and toxins[629].

* except for some of the Biocrates platforms which use direct sampling from plates instead of pre-separation

Consequently, the disruption of the normal balance between the gut microbiota and host (dysbiosis), has been associated with variety of diseases including obesity, malnutrition, inflammatory bowel diseases, neurological disorders, and cancer[635-641]. Studies on the role of dysbiosis in eye diseases are lacking but a very small recent case-control study found that patients with age-related macular degeneration had a significant increase in bacterial gene products linked to L-alanine fermentation, glutamate degradation and arginine biosynthesis pathways and a decrease in the expression of genes of the fatty acid elongation pathway[642]. Abnormalities in glutamate and arginine metabolism have been linked to retinal disease through an ability to promote oxidative stress[643,644], while the role of alanine metabolism is not well studied. The human retina has unique lipid profiles enriched in long-chain and very-long chain polyunsaturated fatty acids that appear to promote normal retinal structure and function[645,646]. Mutations in genes for elongating long-chain polyunsaturated fatty acid cause early onset autosomal dominant retinopathy with similar clinical picture to AMD[645].

The structure of the human microbiome is constantly shifting due to the various environmental (e.g. diet, medication, influx of transient species) and host-related factors (host age and genetics)[631,647]. However, samples obtained over time from the same individual are more similar to one another compared to those obtained from other individuals[472,648,649]. This suggests that each person has a relatively distinct, stable community and there is also the idea of a “core” microbiome representing whatever may be in common among the microbiomes of all humans in terms of functional relationships between the microbiome and the host[647]. Of the environmental factors, diet is perhaps the most significant one for the human gut microbiome, and its role has been supported by both population-based observation studies[650-652] and dietary intervention studies[653,654]. In brief, all macronutrients (protein, fats, digestible and non-digestible carbohydrates), as well as probiotics and polyphenols, can induce shifts in the microbiome with secondary effects on host immunologic and metabolic markers (**Figure 6.2.**). For example, intake of animal protein increases overall microbial diversity (including total anaerobic microflora) as well as the abundance of bile-tolerant organisms (e.g. Bacteroides, Alistipes); and simultaneously reduces representation of the Roseburia/E. rectale group[655]. A diet rich in high-saturated fat results in increase in Actinobacteria and Lactic acid bacteria, while intake of carbohydrates enriches the gut with Bifidobacterium and suppress Clostridia, Lactobacillus, Ruminococcus and others[655].

Figure 6.2: Example of the impact of three common diets on the gut microbiome



This graph is adopted from Singh *et al.*, J Transl Med. 2017[655] and shows the effect of three common diets (Western diet, Mediterranean diet, and Gluten free diet) on intestinal microbiota and on cardio-metabolic traits. CVD — cardiovascular disease, DM2 — type 2 diabetes mellitus.

6.3. Material and Methods

6.3.1. Subjects and phenotyping

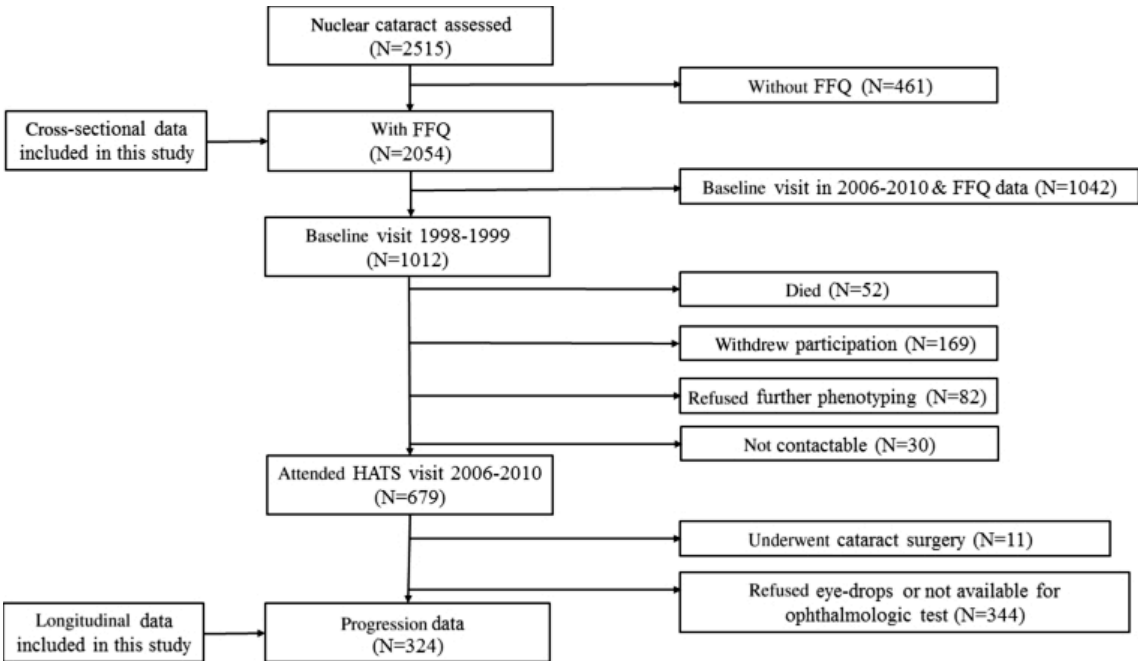
Of the 2,556 twins for whom cataract data was available, only 41 were males and most of those did not complete a FFQ. In addition, only female participants contributed data on cataract progression. Therefore, the male participants were excluded from the analysis presented in this chapter. Nuclear and cortical cataract at baseline (1998-1999, 2006-2010)[167] were measured in 2,515 white female twins (mean age, 62.3 years; range, 50.1–83.1 years) from the TwinsUK cohort, 2,054 of whom had also completed a FFQ around the time of their eye examination (median = 2 years). Nuclear and cortical cataract was measured as detailed in Chapter 4 (4.3.2. Phenotyping). The 461 twins with cataract data but without FFQ data were younger by 2.5 years on average. As a result they were also less affected by cataract.

Cataract progression was measured in 324 twins (151 MZ and 173 DZ pairs) with a mean age at follow-up of 69.8±5.4 years (range, 58.3–83.6 years) as part of the 2006-2010 recruitment period of Healthy Ageing in Twins (HATS) study[464]. The mean time between baseline and second visits was 9.4 years (range, 7–12 years). Nuclear and cortical cataract progression was measured in two ways: a) as a differences: Δscore = score at follow-up – score at baseline; b) as a rate: Δscore = (score at follow-

up – score at baseline)/score at baseline. As both measures yielded the same results, I present the results from the difference score only.

The smaller number of individuals with follow-up data is a result of the fact that the HATS study was not designed specifically as a cataract follow-up study. In HATS, individuals (≥ 40 years of age) who had previously attended clinical phenotyping were invited to participate and this was done irrespective of whether they had an eye examination prior to HATS (**Figure 6.3.**). Thus for 1,523 individuals (aged ≥50 years) the HATS study formed their baseline cataract assessment. Additional reasons for non-participation of the original cohort participants who had cataract measurements in 1998/9 were: death (N = 52), withdrawal of participation from the TwinsUK registry (N = 169), noncontactable (N = 30), refusal of further phenotyping (N = 82), bilateral cataract surgery (N = 11), and refusal of dilating drops or unavailability of ophthalmic testing at the HATS visit (N = 344).

Figure 6.3: Consort digram of the cataract progression part of the study



The figure shows the number of individuals who participated in the different parts of the study and reasons for non-participation at follow-up. FFQ = food frequency questionnaire; HATS = Healthy Ageing in Twins. This figure is adopted from Yonova-Doing *et al.* Ophthalmology. 2016 Jun; 123(6): 1237–1244[656].

6.3.2. Nutrient intake and dietary scores

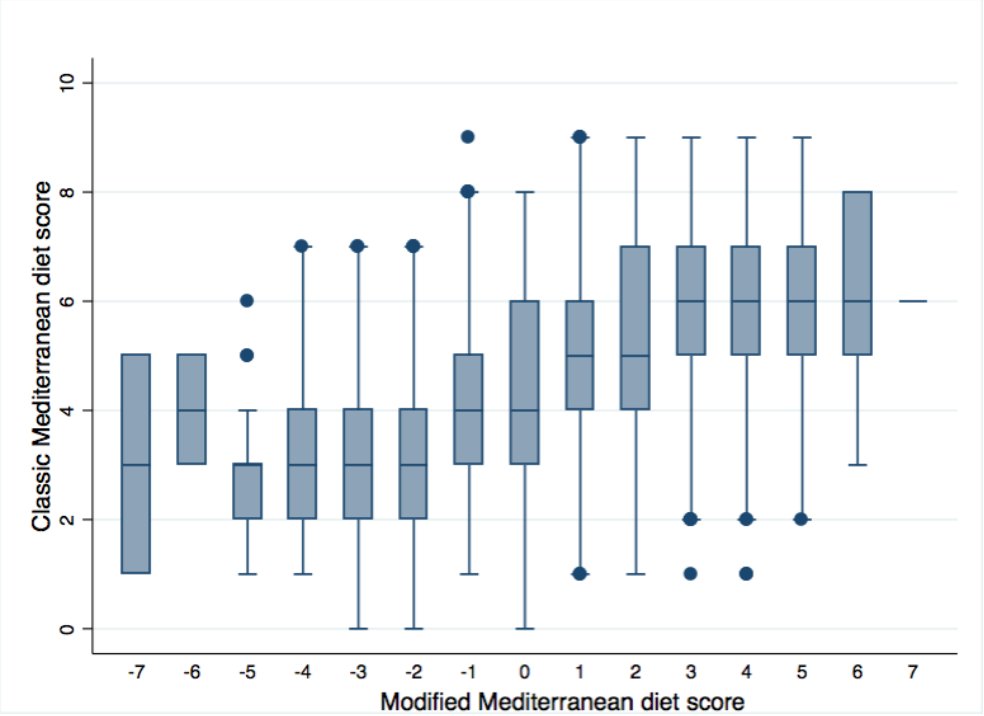
The intake data of nutrients and food groups was originally calculated by by Alex MacGregor and Sue Fairweather-Tait, nutritional epidemiology experts from the University of East Anglia, who have analysed DTR FFQs over 20 years. Briefly, intake of micronutrients (vitamins and minerals) and supplements was estimated at the baseline visit using self-administered EPIC FFQ[657]. This questionnaire explored the average frequency of intake of 131 foods and supplements over a 1-year period[657,658]. Intake per nutrient and per food group was calculated using an established nutrient database[659], and the dietary variables were adjusted for calorie intake, yielding an energy-adjusted mg/µg of each nutrient per person per day[658,660]. The following micronutrients were considered in the analysis: sodium, potassium, calcium, magnesium, phosphorus, iron, copper, zinc, chloride, manganese, iodine, retinol, carotene, vitamin D, vitamin E, thiamine, riboflavin, niacin, tryptophan, vitamin B6, vitamin B12, folate, pantothenate, biotin, and vitamin C. High vitamin K has been found to protect against cataract surgery[415] and is, therefore, another nutrient worth considering. Vitamin K was missing from the original nutrient intake calculations and to obtain data on its intake I used the energy-adjusted food groups data and the latest edition of the nutrient database[661].

Data on supplement intake for 33 different supplements was also available. However, the percentage of individuals taking any single supplement was 10% or less. Supplements were therefore grouped as follows: any supplements; micronutrient supplements (vitamins and minerals in any combination); micronutrient supplements excluding multivitamins (e.g., vitamin C only, vitamin D only, iron only, ACD complex); minerals only (e.g., iron only, calcium only); and other supplements (e.g., aloe vera, Echinacea, Ginkgo, omega-3). Each supplement group was coded as a binary variable, with “yes” indicating that the person took 1 or more of the supplements in a specific group.

There are a lot of different ways to calculate MDS[607]. The alternative MDS was used here[662]. There were two reasons for that choice: the score uses median intake which is slightly more accurate than number of portions; the score does not take into account a component of the Mediterranean diet that was not assessed in TwinsUK, namely olive oil consumption. The score assumes that only people with above median intake are protected but that there is no difference between low and median intake and, therefore, that low intake does not confer any risk. To avoid making such an

assumption, score was modified by giving -1 point to people whose consumption was below the median for every component hypothesised to benefit health (vegetables, fruits, nuts, grains, legumes, fish, ratio of monounsaturated to saturated fat intake) and -1 point was above the median for every component be detrimental to health (red meat, dairy, dairy products). Thus the score used here ranges between -9 and 9. No points were given to people with median intake (**Figure 6.4.**). As opposed the MDS, the HEI, gives a score using all food groups (fruit, vegetables, grains, milk, meat, legumes, oils, sodium, and percentage of energy from saturated fat, alcoholic beverages, and added sugars). In each group different amount of points are given dependent on how much of each food is consumed. The HEI was calculated by Ruth Bowyer as previously described[663].

Figure 6.4: Comparison between the classic and modified MDS



This box plot shows the relationship between the classic and the modified MDS.

6.3.3. Metabolite and microbiome measurements

The methods used to acquire metabolite and microbiome measurements are discussed in detail in Chapter 3 (Section 3.2.2. Omics data).

6.3.4. Statistical analysis

All analyses were carried using the STATA10 or STATA14* statistical packages.

Heritability analysis: The heritability analysis were conducted on 155 twin pairs (72 MZ and 83 DZ pairs) for nuclear and 148 twin pairs (75 MZ and 73 DZ pairs) for cortical cataract using a classical twin modelling implemented in OpenMX (http://openmx.ssri.psu.edu/). In both cases the differences between the follow-up and the base-line measurement was used. In the case of cortical cataract, both people who progressed and people who did not were included in the analysis.

Data transformation: The central nuclear dip score and the nuclear and cortical cataract progression variable were not normally distributed and, therefore, were transformed using natural logarithm before the analysis. The cortical cataract score was treated as binary variable with case-control cut of 5% lens area affected by cortical spokes.

Group comparisons: Means and proportions between groups of individuals (with or without follow-up data, MZ and DZ twins, etc.) in terms of age, cataract scores, and nutritional variables were performed using 2-sample, 2-tailed t-tests or z-tests, assuming equal variance.

Regression analysis:

1. Micronutrient, supplements, dietary scores

A) Initially, the associations between baseline cataract scores and micronutrients or supplement groups were assessed using age-adjusted univariable linear regression models of the kind:

$$y = \alpha + \beta_1 x + \beta_2 age + \varepsilon, \text{ cluster (family\_id) (nuclear cataract)}$$

$$odds(y) = \exp(\alpha + \beta_1 x + \beta_2 age + \varepsilon) \text{ , cluster (family\_id) (cortical cataract)}$$

where: (y) is cataract score; (x) is a predictor of interest (a nutrient or a supplement groups); age is age at baseline; and α is the intercept term. The association between x and y is given by β_1, defined as the amount of change that occurs with one unit change in x. All formulas presented in this section use the same notation as presented here.

All nutrients or supplement groups showing significant univariable association (p-value < 0.05) were then included in a multivariable linear regression model. Independent variables were identified using a stepwise backward procedure with threshold for removal set at 0.05.

---

* StataCorp LP, College Station, TX; www.stata.com

Factors showing significant ($P < 0.05$) association at baseline in the multivariable models were tested for association with progression using the above presented formulas but with adjusted for both age at baseline and Δage (age at follow-up – age at baseline).

The results presented below are for models where the micronutrients were not standardised and were, therefore, on their original scale. Standardisation of nutrients did not yield different results in terms of association (data not shown).

B) In the case of both cross-sectional and progression nuclear cataract scores, the use of transformed scores makes clinical interpretation more difficult. To facilitate the clinical interpretation of the results, relative risk ratios (RRs) were calculated using multinomial regression:

$$\text{RR (category 2 versus reference)}=\exp(\alpha + \beta_1^{(2)}x_1 + \ldots + \beta_i^{(2)}(x_i+1) + \ldots + \beta_k^{(2)}x_k)/$$
$$\exp(\alpha_1 + \beta_1^{(2)}x_1 + \ldots + \beta_i^{(2)}x_i + \ldots + \beta_k^{(2)}x_k), \text{ cluster (family\_id)}$$

$$\text{RR (category 3 versus reference)}=\exp(\alpha + \beta_1^{(3)}x_1 + \ldots + \beta_i^{(3)}(x_i+1) + \ldots + \beta_k^{(3)}x_k)/$$
$$\exp(\alpha_2 + \beta_1^{(3)}x_1 + \ldots + \beta_i^{(3)}x_i + \ldots + \beta_k^{(3)}x_k), \text{ cluster (family\_id)}$$

In this case, the nuclear cataract score, the nuclear cataract progression scores, and the associated nutrients were divided into tertiles, and the first tertile was set as reference. Supplement intake per supplement group was kept binary.

C) Dietary scores:

The following regression models were used to assess the association between the two diet scores and the nuclear and cortical cataract scores respectively were:

$$y = \alpha + \beta_1\text{dietary score} + \beta_2\text{age}_{\text{eye\_visit}} + \beta_3\Delta\text{age} + \varepsilon, \text{ cluster (family\_id)}$$

$$\text{odds}(y) = \exp(\alpha + \beta_1\text{dietary score} + \beta_2\text{age}_{\text{eye\_visit}} + \beta_3\Delta\text{age} + \varepsilon), \text{ cluster (family\_id)}$$

where y in cataract score and Δage = absolute difference between the age at eye visit and the age at which the FFQ was filled in.

2. Metabolites, diets, and cataract

The association between diet scores and single-point metabolite measurements was assessed using the following mixed effects (REML) model:

$$y = \alpha + \beta X + uZ + \varepsilon$$

where: y is the relative concentration of each metabolite; α is the intercept term; β represents the fixed effects of a fixed-effect matrix (X). The following variables were

considered to have fixed effect: dietary score, age, body mass index (BMI), ΔT= age at FFQ response - age at metabolite sampling. Only samples where ΔT ≤ 5 years were considered. The u term represents the random effect of a random-effect matrix (Z). The following predictors were included as having random effect: family id, blood source (serum/plasma), batch and run-day. In order to exclude the possibility that the results were confounded by diabetes-related metabolic changes, an additional model with diabetes status as covariate was also considered.

As technical covariates were accounted for in the initial data normalisation step, the association between diet scores and the mean of the relative metabolite concentrations over several visits was assessed using the following models:

$$y = \alpha + \beta_1 \text{dietary score} + \beta_2 \text{BMI} + \beta_3 \text{age}_{FFQ} + \varepsilon, \text{ cluster (family id)}$$

$$y = \alpha + \beta_1 \text{dietary score} + \beta_2 \text{BMI} + \beta_3 \text{ diabetes status} + \beta_4 \text{ age}_{FFQ} + \varepsilon, \text{ cluster (family id).}$$

Metabolites that were significantly associated with diet scores at p-value less than the Bonferroni threshold (0.05/number of independent tests) for multiple testing were evaluated for association with cataract. As majority of metabolites are correlated, a multiple testing threshold was calculated using the number of effective tests method[664]. This resulted in α=0.0004 (0.05/120*) for single-point metabolites versus diet analysis and in α=0.0003 (0.05/182*) for the mean concentration of the metabolites over time versus diet analysis.

The association between single-point relative metabolite concentrations and cataract was calculated as follows:

$$y = \alpha + \beta X + uZ + \varepsilon \text{ (nuclear cataract)}$$

$$\text{odds}(y) = \exp(\alpha + \beta X + uZ + \varepsilon) \text{ (cortical cataract)}$$

where y is either ln(central nuclear dip score) or cortical cataract case-control status, β represents the fixed effects of a fixed-effect matrix (X). The following variables were considered to have fixed effect: relative metabolite concentration (either or mean concentration), age, BMI, diabetes status, ΔT= age at eye visit - age at metabolite sampling. Only samples where ΔT ≤ 5 years were considered. The u term represents the random effect of a random-effect matrix (Z). The following predictors were included as having random effect: family id, blood source (serum/plasma), batch and run-day.

The association between mean relative metabolite concentrations and nuclear and cortical cataract respectively was calculated as follows:

* number of independent tests

$$y = \alpha + \beta_1 \text{metabolite} + \beta_2 \text{BMI} + \beta_3 \text{age}_{eye\_visit} + \varepsilon, \text{cluster (family id)}$$

$$\text{odds}(y) = \exp(\alpha + \beta_1 \text{metabolite} + \beta_2 \text{BMI} + \beta_3 \text{age}_{eye\_visit} + \varepsilon), \text{cluster (family id)}$$

3. Microbiome, diets, and cataract

The association between relative abundance of OTUs and diet was assessed as follows: $y = \beta_1 \text{dietary score} + \beta_2 \text{BMI} + \beta_4 \text{age}_{FFQ} + \beta_4 \Delta\text{age} + \varepsilon$, cluster (family id), where Δage = age at which the FFQ was filled - age at microbiome sampling (only sample where $\Delta\text{age} \leq 5$ years were considered). The p-value for associations between microbes and diet score were FDR5%-adjusted. All OTUs associated with diet scores were than tested for association with cataract:

$$y = \alpha + \beta_1 \text{OTU} + \beta_2 \text{BMI} + \beta_3 \text{age}_{eye\_visit} + \beta_4 \Delta\text{age} + \varepsilon, \text{cluster (family id)}$$

$$\text{odds}(y) = \exp(\alpha + \beta_1 \text{OTU} + \beta_2 \text{BMI} + \beta_3 \text{age}_{eye\_visit} + \beta_4 \Delta\text{age} + \varepsilon), \text{cluster (family id)}$$

where Δage = age at eye visit - age at microbiome sampling. Only sample where $\Delta\text{age} \leq 5$ years were considered. The Bonferroni threshold was again calculated using number of independent test[664], which resulted in $\alpha = 0.003$ (0.05/17*). Finally, mediation analysis of the relationship between cataract, diet and OTUs were performed using the Sobel-Goodman test and structural equation modelling as implemented in STATA[665]. The mediation analyses were performed in one twin per pair and were adjusted for the effects of age and BMI.

Metabolite enrichment analysis: Metabolite enrichment analysis were carried for all metabolites that were associated (p-value < 0.05) with any of the diet scores or with nuclear cataract using MetaboAnalyst v3.0† and any pathways were considered significantly enriched if their q value (FDR5% adjusted p-value) was < 0.05.

6.4. Results

6.4.1. Micronutrients and the heritability of cataract progression

Baseline characteristics: Baseline characteristics and nutrient and supplement intake for twins with and without follow-up data are presented in **Tables 6.1** to **6.3**. At baseline, the twins with follow-up data (N=324) when compared to the twins without follow-up (N=1730) were one on average a year younger (60.4 vs. 61.5 years, p-value

* number of independent tests

† http://www.metaboanalyst.ca

> 0.05) and had lower nuclear cataract scores (p-value > 0.05; mean nuclear score of 55.3 vs 60.4). Conversely to nuclear cataract, the prevalence of cortical cataract was significantly higher (p-value < 0.0001, 25% vs 15%) in the twins with follow-up data. Finally, the twins with follow-up data consumed on average 0.2mg (p-value = 0.02) less iron and 0.1mg (p-value = 0.04) less thiamine and 0.1mg more biotin (p-value = 0.01).

There were no statistically significant differences (p-value > 0.05) between MZ and DZ twins, except for the MZ twins with cross-sectional data only who had a slightly higher nuclear cataract score (61.6 vs. 59.3; p-value = 0.02). The MZ twins with follow-up data had slightly higher prevalence of cortical cataract (27% vs 23%) but the difference was not statistically significant (p-value = 0.39).

Table 6.1. Baseline sample characteristics in individuals with or without follow-up data

	Subjects without Follow-up			Subjects with Follow-up		
	Total	**MZ**	**DZ**	**Total**	**MZ**	**DZ**
No. of individuals	1730	827	916	324	151	173
Zygosity ratio (MZ:DZ)	01:01.1	-	-	01:01.2	-	-
Age (mean ± SD)	61.5±6.5	61.7±6.7	61.4±6.4	60.4±5.1	60.8±5.5	60.0±5.2
CNDS (mean ± SD)	60.4±17.2	61.3±17.4	59.0±14.2	55.3±11.2	55.3±11.4	55.3±11.1
% with CC	15.7	14.9	16.2	25.0	27.2	23.1

This table presents the baseline characteristics of the participant with and without follow-up data. DZ — dizygotic; MZ — monozygotic; NDS — central nuclear dip score; CC — cortical cataract; *Denotes statistically significant difference (p-value < 0.05) between subjects with and without and without follow-up. The table is adapted from Yonova-Doing *et al* Ophthalmology. 2016 Jun;123(6):1237-44[656].

Heritability of cataract progression: Nuclear cataract scores progressed in all participants (**Figure 6.5**). The mean (SD, [range]) CNDS at baseline was 55 (11, [32-99]) and it increased by 19.9 (16.9, [1–137]) over the period of follow-up. Cortical cataract score progressed in only 32% of participants (**Table 6.4**).

The intraclass correlation coefficient for nuclear cataract progression in MZ and DZ twins was 0.22 and 0.26 respectively. The heritability analysis showed that the best-fitting model for both progression phenotypes was one explained by additive genetic factors and unique environmental factors. The AIC values for the different model were: 1987.77 for ACE; 1985.82 for AE; 1986.46 for CE and 1992.35 for E. In the case of nuclear cataract, 35% [95%CI: 13%-54%] of variance was due to genetic and 65% [95% CI: 46%-87%] of variance was due to unique environmental factors, while the

heritability estimate for cortical cataract progression was 17% [0%-34%], with individual environmental factors explaining 83% of variance [0%-100%].

Table 6.2. Baseline micronutrient intakes in individuals with or without follow-up data

	Subjects without Follow-up			Subjects with Follow-up		
	Total	MZ	DZ	Total	MZ	DZ
Sodium (mg)	2262.8±508.7	2265.3±476.3	2258.7±535.6	2237.4±456.4	2227.7±444.4	2247.2±444.4
Potassium (mg)	4013.5±637.4	3997.0±622.4	4026.9±650.6	4033.7±580.5	4094.5±588.4	3972.5±469.4
Calcium (mg)	1117.1±284.7	1118.5±284.9	1125.1±284.6	1118.9±291.5	1138.3±295.0	1099.4±568.0
Magnesium (mg)	347.3±56.4	347.3±56.8	347.2±56.0	343.8±55.0	347.0±58.0	340.6±287.5
Phosphorus (mg)	1527.1±247.0	1527.1±234.9	1527.1±257.8	1522.0±239.3	1532.0±251.0	1512.1±227.5
Iron (mg)*	13.1±3.0	13.2±3.2	13.0±2.8	12.6±2.6	12.5±2.7	12.7±2.5
Copper (mg)	1.5±0.5	1.5±0.6	1.5±0.4	1.5±0.4	1.5±0.4	1.6±0.5
Zinc (mg)	10.2±1.7	10.2±1.8	10.1±1.7	10.2±1.7	10.2±1.8	10.1±1.6
Chloride (mg)	3629.6±792.9	3633.6±749.4	3623.0±828.6	3578.0±721.3	3566.7±690.3	3589.4±753.3
Manganese (mg)	4.2±1.2	4.1±1.1	4.2±1.2	4.2±1.1	4.3±1.1	4.2±1.1
Iodine (mg)	225.0±75.8	224.2±75.2	225.8±76.5	229.2±64.2	230.0±61.4	228.5±67.2
Retinol (µg)	579.5±817.8	569.1±570.6	554.8±496.6	611.8±472.9	588.2±422.6	635.6±519.0
Carotene (µg)	5343.4±3067.4	5503.7±3263.8	5200.4±2874.9	5305.6±3915.4	5663.8±4823.8	4945.0±2679.4
Vitamin D (µg)	2.7±1.4	2.7±1.1	2.6±1.5	2.8±1.1	3.0±1.0	2.6±1.0
Vitamin E (mg)	11.5±3.2	11.6±3.4	11.4±3.1	11.7±3.4	11.9±3.6	11.5±3.2
Thiamine (mg)*	1.8±0.4	1.8±0.4	1.8±0.4	1.7±0.3	1.7±0.3	1.7±0.3
Riboflavin (mg)	2.5±0.7	2.4±0.7	2.5±0.7	2.4±0.6	2.5±0.6	2.4±0.7
Niacin (mg)	22.0±5.7	22.2±5.1	21.8±6.2	21.3±4.5	21.3±4.6	21.2±4.4
Tryptophan (mg)	17.4±3.0	17.5±2.7	17.3±3.3	17.2±2.5	17.3±2.5	17.1±2.6
Vitamin B6 (mg)	2.6±0.6	2.6±0.6	2.5±0.5	2.5±0.5	2.5±0.5	2.5±0.5
Vitamin B12 (µg)	6.5±3.2	6.7±3.6	6.4±2.9	6.7±2.3	6.7±2.3	6.7±2.4
Folate (µg)	402.2±113.1	400.7±114.0	403.2±112.3	395.7±98.9	402.0±95.9	389.4±101.8
Pantothenate (mg)	7.4±16.0	7.5±21.3	7.2±8.6	6.8±4.2	6.5±2.1	7.1±5.6
Biotin (mg)*	48.1±10.5	47.7±10.3	48.5±10.8	49.7±10.3	50.6±10.2	48.7±10.3
Vitamin C (mg)	165.1±73.9	167.6±74.2	163.0±73.7	166.8±65.0	166.9±68.1	166.7±65.0
Vitamin K (energy-adjusted mg)	20.6±14.4	19.7±13.8	21.6±15.0	21.1±14.7	19.0±10.8	23.2±17.6

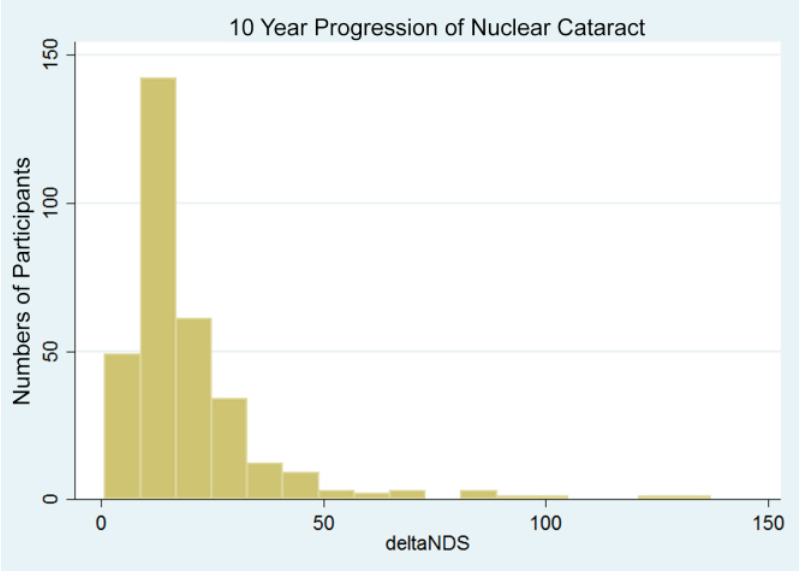
This table presents the micronutrient intakes (mean ± standard deviation (SD))at baseline for people with and without follow-up data. DZ — dizygotic; MZ — monozygotic; *Denotes statistically significant difference (p-value < 0.05) between subjects with and without and without follow-up. The table is adapted from Yonova-Doing *et al* Ophthalmology. 2016 Jun;123(6): 1237-44[656].

Table 6.3. Baseline supplements intakes in individuals with or without follow-up data

	Subjects without Follow-up			Subjects with Follow-up		
	Total	**MZ**	**DZ**	**Total**	**MZ**	**DZ**
Any supplement (%)	55.1	54.8	55.4	55	54.1	55.9
Micronutrients (%)	32.57	32.4	33.2	31.7	32.8	30.8
Micronutrients excluding multivitamins (%)	23.6	24.1	23.2	21.6	24.2	19.3
Minerals only (%)	7.4	7.8	7	6.9	6.4	7.2
Other supplements (%)	44.9	46.2	44.4	47.1	44.2	49.5

This table presents baseline supplements intake per supplement group (% of users) in individuals with and without follow-up data. The supplement groups studied are: any supplement, micronutrient supplements (vitamins and mineral in any combination), micronutrient supplements excluding multivitamins (e.g., vitamin C only, vitamin D only, iron only, ACD complex), minerals only (e.g., iron only, calcium only), and other supplements (e.g., aloe vera, Echinacea, Ginkgo, omega-3). DZ — dizygotic; MZ — monozygotic; *Denotes statistically significant difference (p-value < 0.05) between subjects with and without and without follow-up. The table is adapted from Yonova-Doing *et al* Ophthalmology. 2016 Jun;123(6): 1237-44[656].

Figure 6.5. Progression of nuclear between the 2 visit dates



This figure shows a graphical representation of the progression of nuclear cataract between the 2 visits, presented as deltaNDS (deltaNDS = NDS at follow-up – NDS at baseline) where NDS stands for nuclear dip score. The y-axes show frequency of deltaNDS per bins with width of 6.25 points. The figure is adopted from Yonova-Doing *et al* Ophthalmology. 2016 Jun;123(6): 1237-44[656].

Table 6.4. Cortical cataract progression between baseline and follow-up visits

	Cortical Cataract Score	Baseline	Progression	
		Number of people per score group	Number of people who did not progress	Number of people who progressed
Controls	0	166	166	0
	(0-0.25)	67	28	39
Cases	[0.25-5]	84	22	62

This table shows the number of people whose cortical cataract progressed between the baseline and the follow-up visits, as well as the number of people who showed no measurable progression. The cortical cataract score presented here is the modified Oxford grading system score.

Cross-sectional associations between age-related nuclear cataract and micronutrient intake: Results from univariable cross-sectional regression models are presented in **Table 6.5**. Eight micronutrients (potassium, magnesium, manganese, phosphorus, vitamins C, E and K, and folate) were significantly associated (p-value < 0.05) with nuclear cataract, while only vitamin D was associated with cortical cataract (p-value = 0.04). Only vitamin C (β= –0.0002; SD = 6.3E-05; p-value = 0.01) and manganese (β = –0.009; SD = 0.04; p-value = 0.03) stayed associated with nuclear cataract in the multivariable regression model. Two supplement groups, micronutrient supplements and minerals only, showed a significant (p-value < 0.05) association with nuclear cataract (**Table 6.6.**), but only micronutrient supplements stayed significant in the multivariate model (β=–0.03; SD = 0.01; p-value = 0.01).

The nuclear cataract results from the multinomial regression analysis are presented in **Figure 6.6.**. Here the reduction in relative risk (1 - relative risk ratio (RR)) has been plotted. The RR in this case means the probability that subjects in the highest tertile of nutrient intake would fall in the 2nd or 3rd tertile of CNDS. For example people in the highest tertile of vitamin C intake were 11% less likely to have moderate cataract scores and 19% less likely have high cataract scores. The highest tertile of CNDS encompasses moderate and clinically significant cataracts and, therefore, this result is equivalent to 19% reduction in baseline risk of developing clinically significant cataract.

Associations between age-related nuclear cataract progression and micronutrient intake: Vitamin C was the only nutrient associated with nuclear cataract progression (β = -0.001; SD = 0.001; p-value = 0.03). Individuals in the highest tertile of vitamin C intake had 34% risk reduction of being in the highest tertile of progression.

Table 6.5. Results from univariable regression models between age-related cataracts and micronutrient intakes

	Nuclear cataract			Cortical cataract (358 case/1710 controls)		
	Beta	**SE**	**p-value**	**OR**	**SE**	**p-value**
Sodium (mg)	5.4E-06	9.6E-06	0.56	1.00	1.0E-04	0.34
Potassium (mg)	**–1.6E-05**	**7.5E-06**	**0.04**	1.00	1.0E-04	0.60
Calcium (mg)	–2.0E-05	1.5E-05	0.2	1.00	2.0E-04	0.90
Magnesium (mg)	**–0.010**	**0.004**	**0.01**	1.00	0.001	0.90
Phosphorus (mg)	**–4.0E-05**	**1.9E-05**	**0.04**	1.00	3.0E-04	0.53
Iron (mg)	–1.2E-04	0.002	0.95	1.00	3.0E-04	0.95
Copper (mg)	0.001	0.008	0.86	0.79	0.020	0.09
Zinc (mg)	–7.8E-04	0.003	0.77	0.98	0.107	0.51
Chloride (mg)	3.8E-06	6.1E-06	0.53	1.00	0.037	0.39
Manganese (mg)	**–0.010**	**0.004**	**0.01**	0.97	1.0E-04	0.59
Iodine (mg)	–1.1E-04	6.1E-05	0.07	1.00	0.053	0.21
Retinol (µg)	2.4E-06	3.9E-06	0.55	1.00	0.001	0.94
Carotene (µg)	–1.7E-06	1.4E-06	0.23	1.00	2.0E-05	0.40
Vitamin D (µg)	–0.004	0.003	0.22	**1.11**	**0.040**	**0.01**
Vitamin E (mg)	**–0.003**	**0.001**	**0.04**	1.01	0.019	0.77
Thiamine (mg)	–0.013	0.013	0.3	0.91	0.152	0.58
Riboflavin (mg)	–0.011	0.006	0.08	0.96	0.087	0.64
Niacin (mg)	–1.1E-04	8.3E-04	0.89	1.00	0.012	0.87
Tryptophan (mg)	–0.001	0.001	0.27	1.02	0.195	0.45
Vitamin B6 (mg)	–0.002	0.009	0.81	0.97	0.109	0.81
Vitamin B12 (µg)	–0.001	0.001	0.5	1.00	0.021	0.68
Folate (µg)	**–9.9E-05**	**4.1E-05**	**0.02**	1.00	0.001	0.27
Pantothenate (mg)	–2.8E-05	1.9E-04	0.88	0.98	0.010	0.11
Biotin (mg)	–3.0E-04	4.2E-04	0.47	1.00	0.006	0.51
Vitamin C (mg)	**–1.7E-04**	**6.2E-05**	**0.01**	1.00	0.001	0.55
Vitamin K (mg)	**-0.001**	**3.0E-04**	**0.03**	0.99	0.005	0.19

This table, adapted from Yonova-Doing *et al*. Ophthalmology. 2016 Jun;123(6):1237-44[656], shows the results of the univariable linear regression analysis between the nuclear dip score or binarised cortical cataract score (cut-off: 5% of area with opacity) and energy-adjusted micronutrient intakes. Statistically significant associations (p-value < 0.05) are denoted in bold. Analyses were adjusted for age and family structure. SE – standard error; OR – odds ratio.

Excluding the subjects with the greatest progression (>100 units of change) did not alter the results. Vitamin D, the only nutrient associated with cortical cataract score at baseline, was not associated with progression (p-value = 0.32).

Table 6.6. Results from univariable regression models between cataract and supplement intakes

	Nuclear cataract			Cortical cataract (358 case/1710 controls)		
	Beta	**SE**	**p-value**	**OR**	**SE**	**p-value**
Any supplement	–0.015	0.009	0.12	1.20	0.194	0.27
Micronutrients	**–0.032**	**0.013**	**0.01**	1.21	0.211	0.27
Micronutrients excluding multivitamins	–0.023	0.012	0.06	1.19	0.195	0.29
Minerals only	**–0.038**	**0.016**	**0.02**	0.96	0.232	0.86
Any other supplement	0.005	0.014	0.72	0.90	0.205	0.65

This table shows the results of the univariable linear regression analysis between the natural logarithm-transformed central nuclear dip score or binarised cortical cataract score (cut-off: 5% of area presenting with opacity) and supplement intakes per supplement group. Statistically significant associations (p-value < 0.05) are denoted in bold. † supplement intakes per supplement groups were coded as binary traits (0/1). All analyses were adjusted for age and family structure. SE – standard error; OR – odds ratio. The table is adapted from Yonova-Doing *et al*. Ophthalmology. 2016 Jun;123(6):1237-44[656].

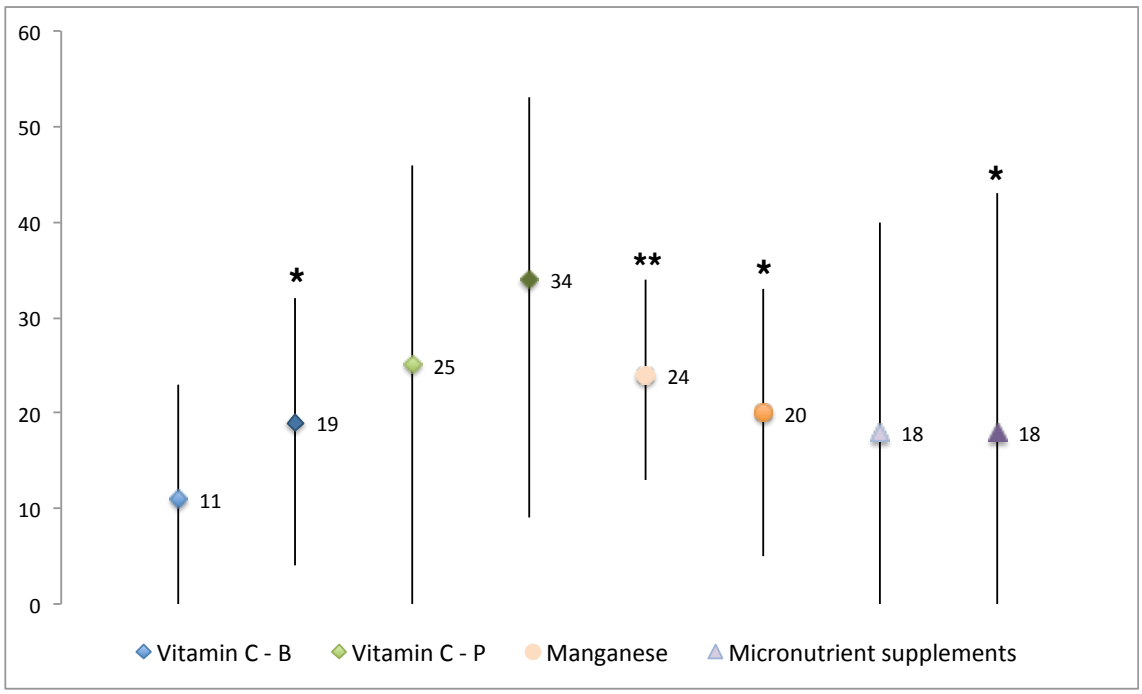
6.4.2. Diets, cataract and the metabolome

The numbers of samples with ARC measurements were much smaller than the number of measured metabolites or OTUs resulting in a multiple testing problem. For this reason the following strategy was used. The association between diet scores and metabolites and between diet scores and microbiome was explored in everyone who had FFQ data (2/3 FFQ questioners)*: cross-sectional metabolite measurements (N = 3000), longitudinal metabolite measurements (N = 2049); microbiome data (N = 1514). Only the associated variants after these analyses were tested for association with ARC in the number of individuals described above. Of the 2054 individuals who had both

Not all metabolites or OTUs were present in all individuals. Metabolites or OTUs were considered only when the sample size for those was more than 100 individuals.

ARC scores and FFQ data: all had metabolites measured crosse-sectionally; 1835 had metabolite measurements at multiple times; and 757 had their gut microbiome sequenced.

Figure 6.6. Relative risk reduction of nuclear cataract risk

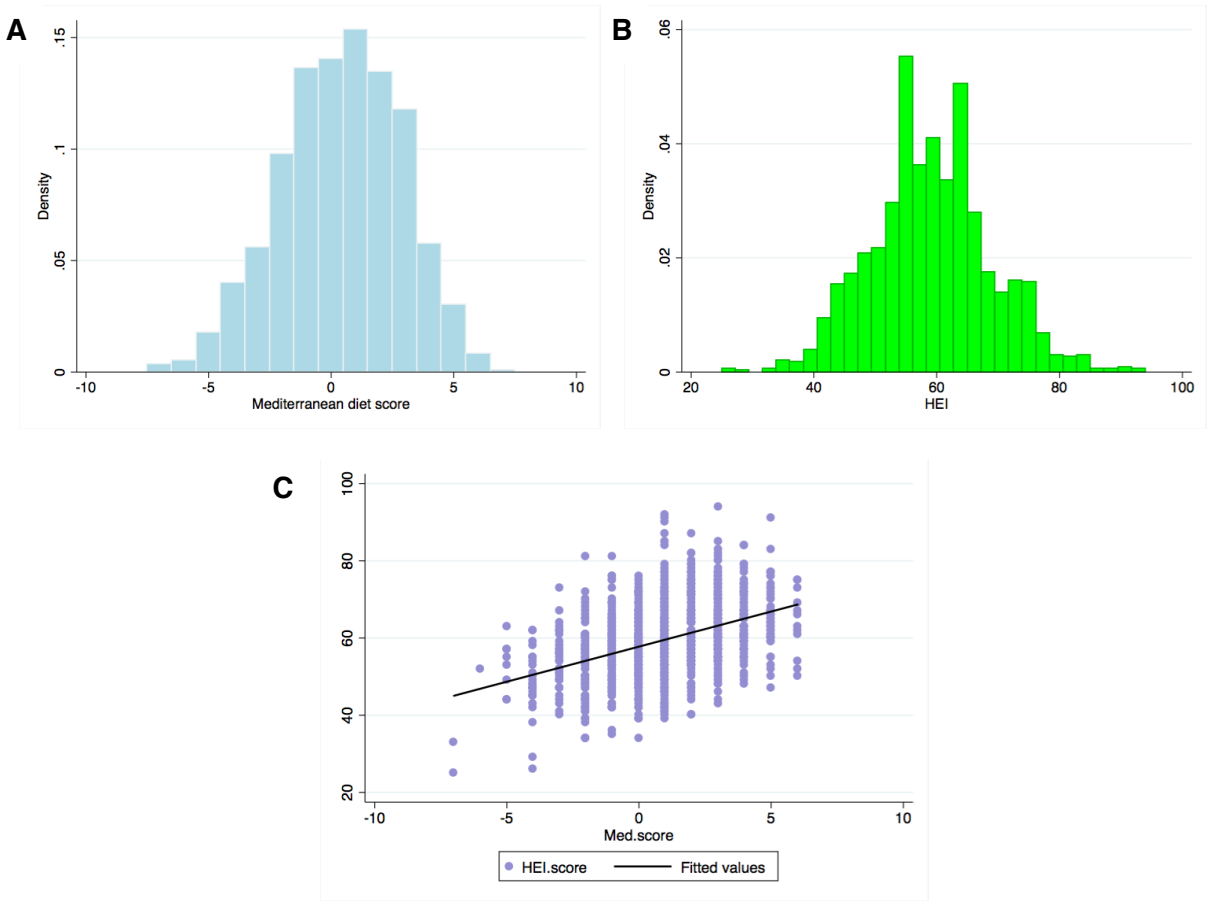


This figure shows the results from the multinomial regression analysis. Relative risk reduction is plotted as a percentage. Diamonds denote vitamin C at baseline (C-B) and in relation to progression (C-P). Circles denote manganese. Triangles denote micronutrient supplements. Lighter colours and darker colours denote - 2nd and 3rd tertile of central nuclear dip score (CNDS) respectively. * — p-value < 0.01; ** — p-value < 0.001. Baseline CNDS tertiles: 1st [34.5-53.2] — reference; 2nd — [53.3-54.5]; 3rd — [54.6-229.2]. ΔCNDS tertiles: 1st [1.0-12.6] — reference; 2nd — [12.7-19.3]; 3rd — [19.4-137.1].

Distribution plots for both scores, as well as a scatter plot showing the relationship between the two scores are presented below (**Figure 6.7.**). In this analysis the median MDS score was 1 (IQR: 3) and the mean (SD) HEI score was 59.1 (9.8). Both dietary scores were associated with nuclear cataract, independently of age and higher index score was associated with lower cataract score (β (se), p-value): -0.006 (0.002), 0.01 for the MDS; and -0.003 (0.001) for the HEI. Only the HEI score was associated with cortical cataract but with a small negative effect OR=1.06, p-value = 0.02. The single point metabolite analysis considered 274 metabolites and of those 71 were associated with the Mediterranean diet and 38 were associated with the HEI but

the sets of metabolites associated with the two dietary scores largely did not overlap. The following pathways were nominally enriched within the MDS association: protein biosynthesis (p-value = 0.01), α-linolenic and linoleic acid metabolism (p-value = 0.02), aspartate metabolism (p-value = 0.04). Only one pathway, glycerolipid metabolism was nominally enriched (p-value = 0.02) within the HEI associations. None of these pathways survived correction for multiple testing (q-value=1.0 for all).

Figure 6.7. Distribution plots for the MDS and HEI scores and a scatter plot showing the correlation between the two scores



This plots show: the distribution of the Mediterranean diet score (MDS, A) and the Healthy Eating Index (HEI, B) score; the correlation between the two score (C).

After correcting for multiple testing, only 20 MDS-associated and 3 HEI-associated metabolites were still significantly associated with cataract (p-value < 0.0004, **Table 6.7**). Diabetes status did not affect the association between diet scores and metabolites, except for the association between MDS and DHA and the association between HEI and 1,5-AG, where where adjusting for it led to smaller p-values.

Table 6.7. Results of the single-point metabolite analysis and nuclear cataract

A) Diets and Metabolites								B) metabolites and nuclear cataract		
Mediteranean Diet				HEI						
beta	se	p-value	p-value2	beta	se	p-value	p-value2	beta	se	p-value
pipecolate (amino acid)										
0.01	**0.002**	**1.0E-04**	**4.0E-05**	-	-	-	-	**-0.07**	**0.02**	**0.005**
glycerate (carbohydrate)										
0.01	0.002	6.3E-08	2.7E-07	-	-	-	-	-0.02	0.04	0.647
eicosapentaenoate (EPA) 20:5n3 (lipid)										
0.01	0.002	1.5E-12	9.1E-11	0.003	0.001	8.0E-04	5.0E-03	-0.03	0.03	0.250
proline (amino acid)										
-4.0E-03	**0.001**	**1.0E-04**	**2.0E-04**	-	-	-	-	**0.13**	**0.05**	**0.016**
docosahexaenoate (DHA) 22:6n3 (lipid)										
0.02	0.002	8.0E-09	8.9E-16	-	-	-	-	-0.04	0.03	0.20
1,5-Anhydroglucitol (1,5-AG, carbohydrate)										
-0.01	0.002	5.0E-05	5.0E-05	-0.002	0.001	7.0E-03	1.0E-05	0.05	0.04	0.20
threonate (co-factors/vitamins)										
0.01	0.002	2.0E-04	5.8E-05	-0.001	0.002	1.0E-02	1.0E-02	-0.02	0.03	0.41
keto leucine (amino acid)										
0.001	0.001	1.6E-06	4.6E-06	-	-	-	-	-0.01	0.05	0.76
3-carboxy-4-methyl-5-propyl-2-furanpropanoate (carbohydrate)										
0.04	0.004	2.2E-12	2.2E-12	-	-	-	-	-0.01	0.01	0.18
2-amino butyrate (amino acid)										
0.01	0.001	4.4E-05	2.7E-05	0.01	0.002	1.5E-07	1.5E-06	-0.09	0.05	0.06
n3 docosapentaenoate (DPA) 22:5n3 (lipid)										
0.01	0.002	7.9E-07	1.1E-05	0.001	4.0E-04	5.0E-03	3.0E-03	-0.02	0.03	0.48
X - 02269										
0.03	**0.004**	**2.6E-12**	**9.0E-12**	-	-	-	-	**-0.04**	**0.02**	**0.03**
X - 11299										
0.03	0.007	1.7E-04	3.9E-04	0.01	0.002	1.3E-07	1.4E-06	-0.01	0.01	0.69
X - 11315										
0.01	0.002	1.6E-07	5.3E-07	-	-	-	-	-0.01	0.04	0.88
3-dehydrocarnitine* (lipid)										
-0.01	0.001	6.6E-05	1.8E-04	-	-	-	-	-0.04	0.05	0.36
C-glycosyl tryptophan* (amino acid)										
-3.0E-03	0.001	4.9E-05	3.7E-05	-0.001	3.0E-04	1.9E-02	1.9E-02	0.04	0.07	0.54
X - 11372										
-0.01	0.002	5.5E-05	9.3E-05	-	-	-	-	-0.02	0.04	0.57
X - 11469										
0.03	0.004	7.0E-11	3.3E-10	0.01	0.002	3.7E-08	4.8E-07	-0.04	0.02	0.05
1-docosahexaenoylglycerophosphocholine (lipid)										
0.01	0.002	1.6E-06	6.1E-06	0.002	0.001	2.0E-03	2.0E-03	-0.04	0.03	0.16
1-oleoylglycerophosphoethanolamine (lipid)										
-0.01	0.002	9.9E-05	2.9E-05	-0.002	0.001	4.0E-03	4.0E-03	0.01	0.03	0.71

This table shows the linear mixed effect model results (beta, standard error (SE), p-value and p-value with diabetes status as covariate (p-value2)) from the diet scores ~ metabolite (A) and nuclear cataract ~ metabolite (B) models. HEI result with p-value > 0.05 are omitted. HEI results

surviving correction for multiple testing are underlined; Metabolites associated with nuclear cataract are denoted in bold; X denotes metabolites quantified by Metabolon which are currently chemically unidentified.

Of the metabolites associated with the MDS, three were associated with nuclear cataract (**Table 6.7.**), and one was associated with cortical cataract (glycerate, OR(SE)=3.33 (1.68); p-value=0.02). Association between HEI-associated metabolites and nuclear or cortical cataract were not observed. Given these results, I next checked if any of the measured metabolites, irrespective of whether they were associated with diet, were associated with nuclear or cortical cataract. Fourteen additional metabolites were associated with nuclear cataract but the strongest association remained the one with pipecolate. No additional metabolites were found to be associated with cortical cataract (data not shown).

The single-point metabolite analysis assumes stable metabolite concentrations over time when in reality metabolite concentrations are quite variable. A way to deal with this is attempting to calculate a steady state concentration, reflecting an average concentration over several visits. In this way possible effects of measurement error and diurnal/seasonal/other variation are reduced. The analysis of the mean metabolite concentration over time considered 738 metabolites* (182 independent metabolites, $\alpha^{\dagger}$ = 0.0003). Of those 199 were associated with the MDS but only 74 survived Bonferroni correction for multiple testing: 26 lipids, 13 Xenobiotics, 6 amino-acids, 2 carbohydrates, 6 vitamins/cofactors, 1 energy-related metabolite, and 20 unknown metabolites. One pathway, protein biosynthesis, was statistically significantly enriched ($p\text{-value}_{adjusted} = 1.9 \times 10^{-4}$) within the 199 metabolites associated with the MDS. Finally, 101 metabolites were associated with the HEI but only 21 survived correction for multiple testing and no pathways were found to be enriched. The associated metabolites (p-value < 0.0003 for at least one of the scores) overlapping between the two diet scores are presented in **Table 6.8.**. Similarly to the single-point analysis, there was no strong evidence for association between metabolites, irrespective of their association with diet, and nuclear or cortical cataract. Adjusting for diabetes status did not significantly change the results of any of the models.

---

* the increase in number of metabolites is duo to fact that a more comprehensive Metabolon platform was used to assay the samples included in the longitudinal study. This platform gives more information on lipid subspecies.

† multiple testing threshold

Table 6.8. Results of the metabolite analysis using the mean metabolite concentration over several visits

Mediteranean Diet				HEI				metabolites & cataract		
beta	se	p	p2	beta	se	p	p2	beta	se	p
ergothioneine (xenobiotics)										
0.09	0.01	7.6E-15	1.8E-13	0.02	0.004	2.6E-07	3.3E-06	-0.001	0.01	0.95
pipecolate (amino acid)										
0.08	**0.01**	**1.3E-13**	**2.0E-11**	**0.02**	**0.004**	**2.7E-08**	**2.4E-07**	0.01	0.01	0.34
1-palmitoyl-2-docosahexaenoyl-GPC (16:0/22:6) (lipid)										
0.08	0.01	2.1E-13	6.2E-13	0.02	0.004	4.5E-08	1.3E-07	-0.02	0.01	0.83
docosahexaenoate (DHA; 22:6n3) (lipid)										
0.08	0.01	1.3E-12	2.5E-11	0.02	0.003	3.9E-06	2.9E-05	0.01	0.01	0.44
oxalate (ethanedioate, co-factors/vitamins										
0.07	0.01	1.6E-11	3.2E-11	0.01	0.003	6.5E-06	4.3E-06	0.004	0.01	0.67
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPE (P-16:0/22:6)* (lipid)										
0.07	0.01	6.1E-10	4.5E-09	0.02	0.004	3.7E-05	2.3E-04	0.001	0.01	0.97
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)* (lipid)										
0.08	0.01	1.7E-09	1.7E-08	0.02	0.004	4.1E-06	1.4E-05	0.002	0.01	0.80
4-allylphenol sulfate (xenobiotic)										
0.06	0.01	3.8E-09	1.1E-08	0.01	0.003	1.9E-06	7.0E-06	-0.14	0.01	0.22
1-palmitoyl-2-eicosapentaenoyl-GPC (16:0/20:5)* (lipid)										
0.06	0.01	4.6E-09	4.3E-08	0.01	0.003	1.0E-05	4.7E-05	0.002	0.01	0.80
1-docosahexaenoyl-GPC (22:6)* (lipid)										
0.07	0.01	9.0E-09	2.0E-08	0.02	0.003	2.3E-07	7.5E-07	-0.001	0.01	0.61
eicosapentaenoate (EPA) 20:5n3 (lipid)										
0.06	**0.01**	**3.4E-08**	**6.7E-07**	**0.01**	**0.003**	**4.9E-05**	**4.2E-04**	0.002	0.01	0.84
X - 23765										
0.06	0.01	6.2E-08	1.4E-06	0.01	0.004	2.7E-05	5.2E-05	-0.003	0.01	0.67
1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6, lipid)										
0.06	0.01	1.0E-07	4.1E-08	0.02	0.004	7.9E-06	8.4E-06	0.003	0.01	0.67
1-docosahexaenoyl-GPE (22:6)* (lipid)										
0.06	0.01	2.6E-07	4.8E-07	0.01	0.003	8.0E-05	2.9E-04	0.01	0.01	0.34
1-eicosapentaenoyl-GPC (20:5)* (lipid)										
0.05	0.01	7.7E-07	6.0E-06	0.01	0.003	2.1E-06	1.4E-05	0.01	0.01	0.34
1-oleoyl-2-docosahexaenoyl-GPC (18:1/22:6)* (lipid)										
0.05	0.01	1.8E-06	2.0E-06	0.02	0.004	1.1E-05	6.1E-06	0.01	0.01	0.06
X - 11315										
0.05	0.01	1.5E-05	1.5E-04	0.02	0.004	1.2E-06	3.8E-06	0.01	0.01	0.48
phosphocholine (18:0/20:5, lipid)										
0.05	0.01	2.2E-05	8.7E-05	0.01	0.003	4.6E-05	1.6E-04	-0.003	0.01	0.75
1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPC (P-18:0/22:6)* (lipid)										
0.05	0.01	2.8E-05	6.4E-05	0.01	0.004	2.5E-04	1.0E-03	0.003	0.01	0.68

This table shows the linear model results (beta, standard error (SE), p-value and p-value with diabetes status as covariate (p-value2)) from the diet scores ~ mean metabolite concentration (A) and nuclear cataract ~ mean metabolite concentration (B) models. HEI result with p-value > 0.05 are omitted. HEI results surviving correction for multiple testing are underlined; Metabolites

associated with nuclear cataract are denoted in bold; X denotes metabolites quantified by Metabolon which are currently chemically unidentified.

6.4.3. Diets, cataract and the microbiome

Of the 560 taxonomy collapsed OTUs, 20 OTUs were associated with the MDS and 60 OTUs were associated with HEI (**Table 6.9.**, **Figures 6.8. and Figure 6.9**). HEI associations were observed within all bacteria classes called in the sample, while Mediterranean diet was associated almost exclusively with changes in the clostridia class of bacteria.

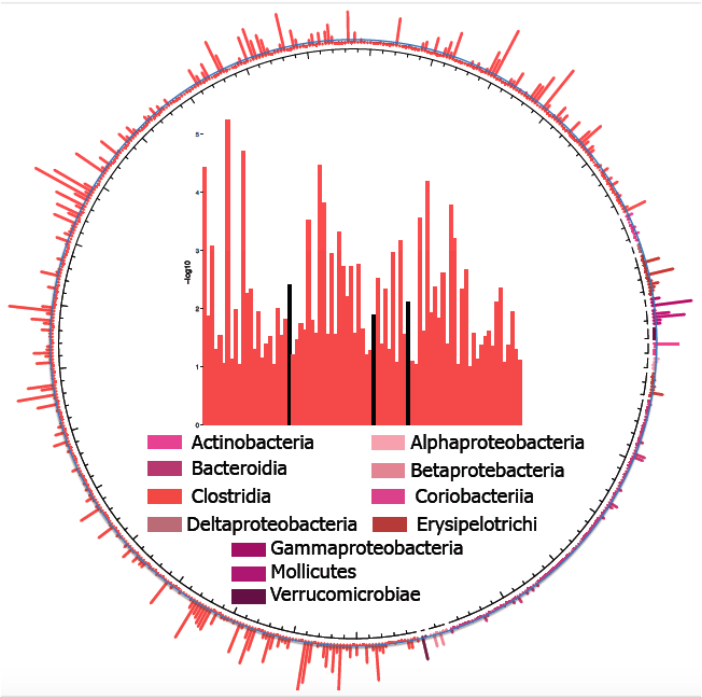
None of the MDS-associated OTUs were associated with nuclear cataract (p-value > 0.05). Only three of the HEI associated OTUs, two belonging to the Ruminococcaceae family (β(SE) = -0.002(0.01) p-value = 0.004 and β(SE) = -0.002(0.01) p-value = 0.02 respectively) and one belonging to the Mogibacteriaceae family (β(SE) = 0.002(0.01); p-value = 0.01) were associated with nuclear cataract. None of the HEI-assocaited bacteria were associated with cortical cataract. No OTUs were found to be associated with cortical cataract (p-value > 0.05). Finally, alpha-diversity (the mean bacterial species diversity per individual) was positively associated with both dietary scores (p-value = 0.01 and p-value = 0.003 for the MDS and HEI respectively) but not with any of the cataract scores. Finally, mediation analysis were also attempted. The results of those are presented in **Figure 6.9.** There was no evidence for mediation by both OTUs when using the Sobel-Goodman test but he structural equation modelling test suggested a mediation effect for both OTU (p-value = 0.049 and p-value = 0.024 for Ruminococcaceae and Mogibacteriaceae OTUs respectively.)

Table 6.9. Associations between diet scores and operational taxonomy units (OTUs)

	beta	se	adjusted p-value
top 10 Mediterranean diet associated OTUS			
o__Clostridiales; f__; g__; s__	0.05	0.012	2.7E-05
o__Clostridiales; f__; g__; s__	**0.05**	**0.013**	**2.7E-05**
o__Clostridiales; f__Ruminococcaceae; g__; s__	0.05	0.012	1.7E-05
o__Clostridiales; f__Ruminococcaceae; g__; s__	0.05	0.012	7.1E-05
o__Clostridiales; f__Ruminococcaceae; g__; s__	0.05	0.012	6.4E-05
o__RF32; f__; g__; s__	0.05	0.012	1.4E-04
o__Clostridiales; f__Ruminococcaceae; g__Faecalibacterium; s__prausnitzii	0.04	0.012	1.4E-04
o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__	0.04	0.012	1.4E-04
o__Desulfovibrionales; f__Desulfovibrionaceae; g__; s__	0.04	0.011	1.8E-04
o__Clostridiales; f__Ruminococcaceae; g__; s__	0.04	0.012	1.7E-04
top 10 HEI associated OTUs			
o__Clostridiales; f__Lachnospiraceae; g__Lachnospira; s__	0.14	0.012	5.6E-06
o__Clostridiales; f__; g__; s__	0.13	0.013	2.0E-05
o__Clostridiales; f__; g__; s__	**0.14**	**0.013**	**3.4E-05**
o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__	0.13	0.014	3.6E-05
o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__	0.12	0.012	6.4E-05
o__Clostridiales; f__Lachnospiraceae; g__[Ruminococcus]; s__gnavus	-0.12	0.013	1.5E-04
o__Clostridiales; f__Lachnospiraceae; g__Lachnospira; s__	0.11	0.013	1.7E-04
o__Clostridiales; f__[Mogibacteriaceae]; g__; s__	-0.13	0.013	2.7E-04
o__Clostridiales; f__; g__; s__	0.11	0.013	3.0E-04
o__Clostridiales; f__; g__; s__	0.11	0.013	3.0E-04

This table presents the 10 strongest association between diet scores (Mediterranean diet score and Healthy eating index score (HEI)) and operational taxonomy units (OTUs), resulting from the linear regression analysis (beta, its standard error (SE), and adjusted p-value — FDR of 5% adjusted p-value). o — order; f — family; g — genre; s — species; Bold text denotes cases where the same OTU was associated with both scores.

Figure 6.8. Associations between HEI and OTUs per bacterial class



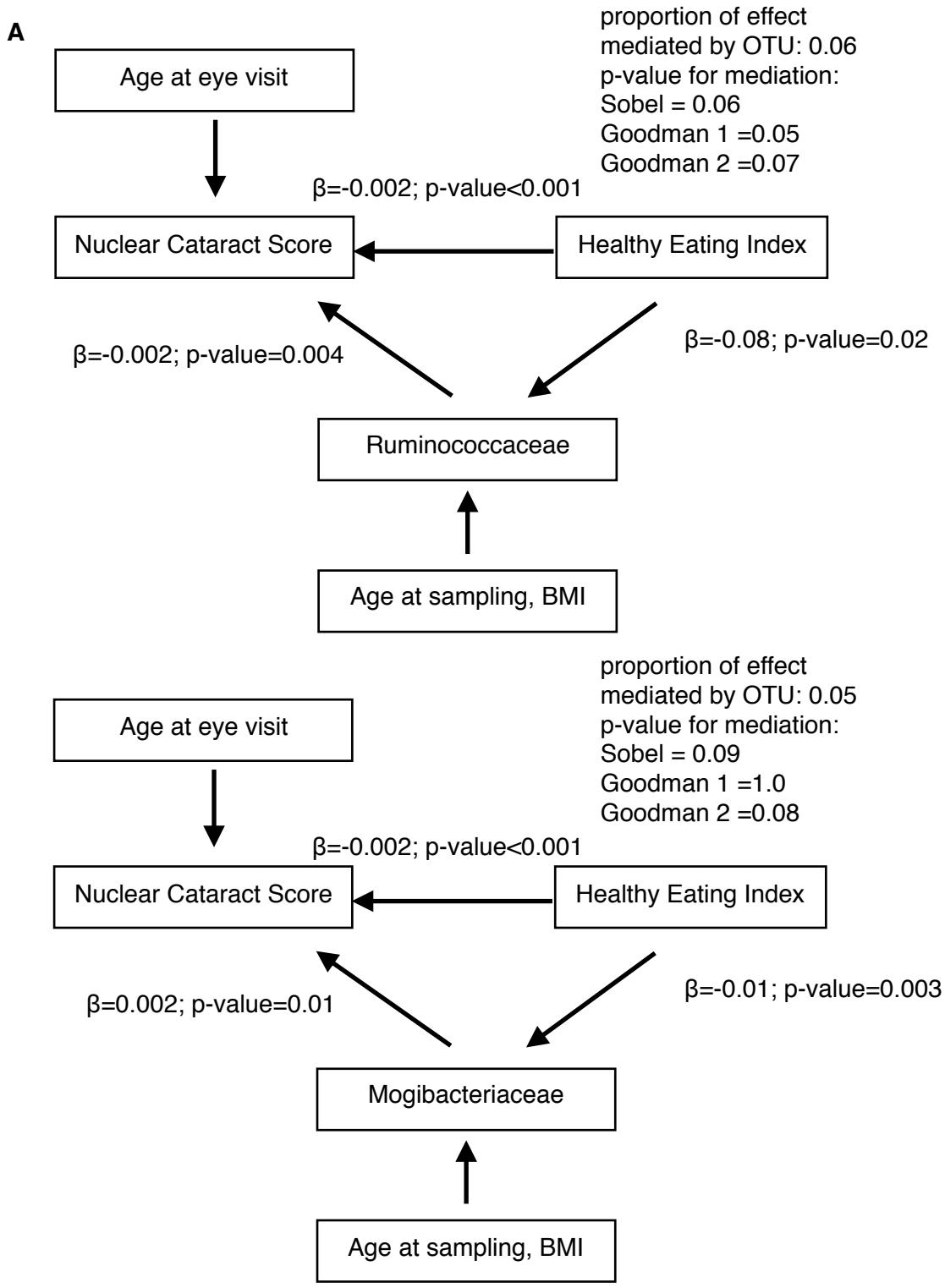
This figure shows the associations between Healthy Eating Index (HEI) and operation taxonomy units (OTUs) per bacterial class (outer ring). The inset shows a Manhattan plot of the association between Clostridia OTUs and nuclear cataract (black) in comparison with all other tested OTUs

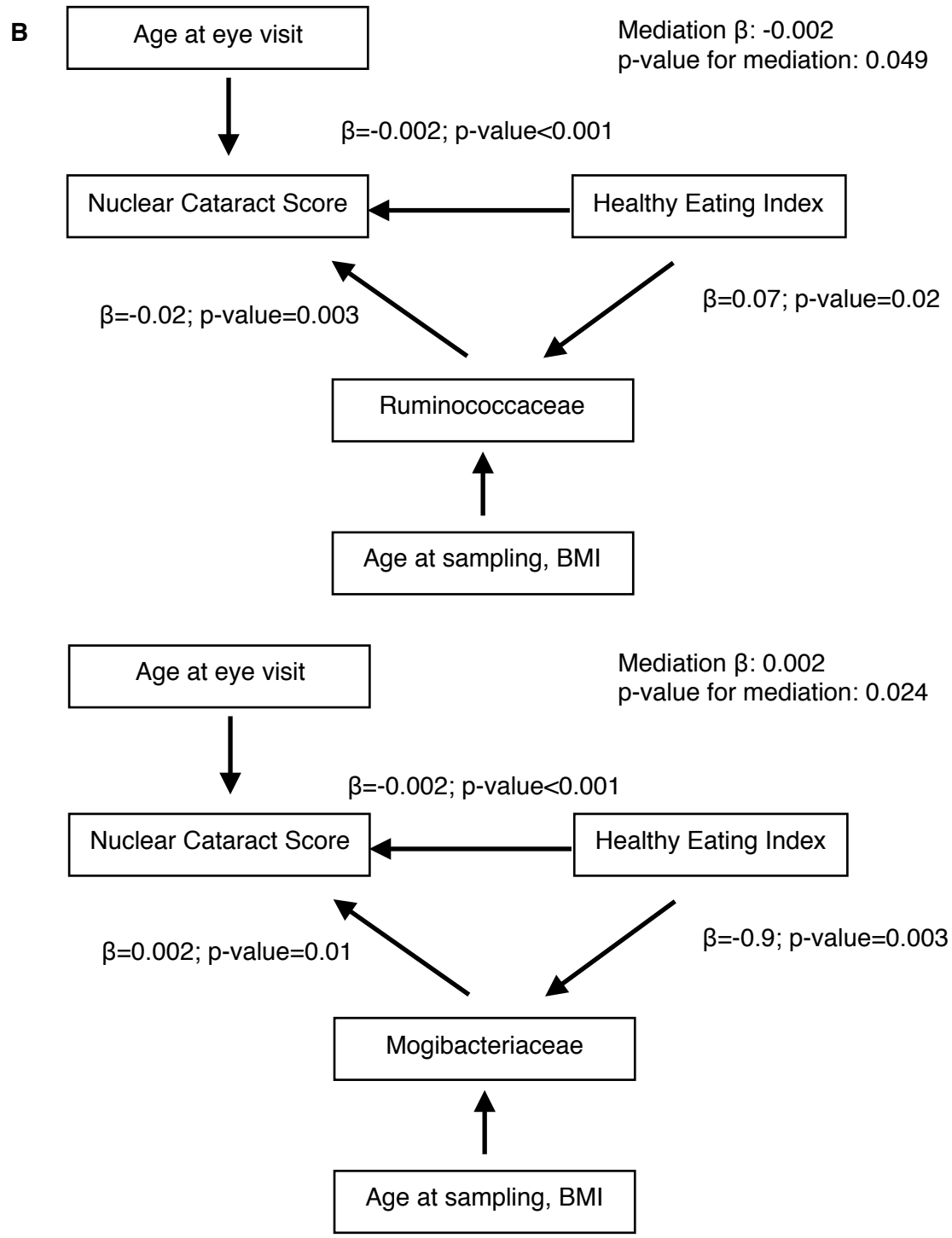
6.5. Discussion

In order to elucidate the relationship between diet and the two most common types of cataract, the following questions were addressed in this chapter: 1) What proportion of variance in cataract progression is explained by genetic factors?; 2) Do micronutrients play a role in ARC progression?; 3) Is the association between diet and the ARC mediated by changes in metabolite concentration? 4) Is the association between diet and the ARC mediated by changes in the gut microbiome?

The results revealed that the progression of nuclear and cortical cataract over a 10-year period in a group of UK female twins was influenced primarily by environmental factors with genetic factors explaining just 35% of 17% of variance respectively. The heritability estimates of cataract progression were lower than previous cross-sectional estimates published for the same population[167,169]. This was not due to some sort of selection bias for those 324 subjects in the follow-up group who contributed the

Figure 6.9. Mediation analysis of diet, cataract and HEI associated OTUs.





This figure presents the results from the mediation analysis exploring whether the effect of diet on cataract is mediated by operational taxonomy units (OTUs). In this analysis only one twin per pair was included. Betas and p-values from Sobel-Goodman test (A) and structural equation modelling (B) are presented for the separate paths and the whole model. In both cases the analyses were adjusted for the effects of age and BMI.

progression data as these estimates are half as large as the baseline heritability

estimates that were recalculated in this group (61% [95%CI: 45%-72%] for nuclear and 50% [95%CI:35-65] for cortical cataract respectively). These findings were, however, consistent with previous studies showing that heritability generally is lower when examining change, compared with cross-sectional studies[666-668]. A variety of other factors might play a role in determining change during ageing than genetic factors. For example, environmentally driven differences in response to stressors in early development or in adulthood may be important, as well age-related accumulation of somatic mutation or epigenetic remodelling[668].

The heritability findings are also dependent the ability of the grading methods used here to detect change. In the case of nuclear cataract a highly quantitative and reproducible* measure of cataract from digital images was used. The fact that every subject measured showed progression suggested that the central nuclear dip score is sensitive to change. CNDS may also be better in capturing change than the widely used LOCSIII system that was developed to increase steps between scores to allow greater sensitivity to change, accepting a lower inter-grader reproducibility[36]. Longitudinal studies using the LOCS III scale show relatively little change. For example, in the Longitudinal Study of Cataract, only 24% of participants had an increase in nuclear opacities over an average of 4.6 years[385]. Although CNDS is not the same measure, it is highly correlated with average nuclear opalescence graded digitally or using a slit lamp[167].

Regarding cortical cataract, 32% of participants progressed in 10-years, thus with 324 participants the study was underpowered to estimate heritability. If everyone would have progressed, 100 pairs would have power of 80% to detect a heritability of 20%. This rate of progression observed here is comparable to those found in other studies, namely close to twice the 5-year progression found in BAES and MVIP and close but slightly higher than the 10-year progression of 20% found in BMES[232,233,669]. Furthermore, none of the participants with grade of 0 at base line developed cortical spokes at follow-up. This implies that women, who by the age of 60 have not developed cortical cataract, are less likely to ever develop cortical cataract. Therefore, for case-control analysis, selecting older controls could prevent misclassification by labelling as controls people who are likely to develop cataract in the future.

Next the role micronutrients play in cataract progression was explored. Those participants who contributed follow-up data (N = 324) were seen as part of the HATS

* intraclass correlation coefficient for the worse eye in 30 subjects who came for a repeat measurement as part of the original nuclear cataract heritability study (Hammond *et al*, N Engl J Med. 2000 Jun 15;342(24):1786-90) was 0.93

study, which was not designed as a cataract follow-up study. This reduced the amount of data available for analysis and, therefore, the power of the study. For that reason, in the single nutrient analysis I focused on micronutrients only. According to the available literature discussed in detail in **Chapter I**, those were more likely to yield strong signal that, if present, could be detected in the progression analysis. Furthermore, only nutrients that were associated with nuclear cataract score at baseline were tested for association with progression.

Vitamin C was found to be inversely associated with nuclear cataract progression; and the previous findings in other population cohorts of association between cross-sectional cataract and vitamin C intake were replicated. Vitamin C intake has been extensively studied in relation to ARC because it is the L-enantiomer of ascorbate — one of the metabolites with the highest concentration in aqueous humor and the lens that reduces oxidation products[670-672]. However, the conclusions of the many studies of the effects of ascorbate on cataract development are inconsistent and often conflicting[272,375-381,387]. Many of these have studied relatively well-nourished populations and are cross-sectional. Cross-sectional studies in India, where overall antioxidant levels may be lower, have found an inverse relationship between vitamin C and cataract[378,409]. These results are in agreement with the CAREDS* and BMES† studies where dietary vitamin C was inversely associated with 10-years prevalence and incidence respectively[377,387]. Dietary manganese was found to be protective against cross-sectional nuclear cataract independent of vitamin C. This association may be a type I error, given that there was no association between dietary manganese and nuclear cataract progression and the lack of a dose response (**Figure 6.6.**), although factors associated with incidence and progression do not always overlap. Manganese is an important antioxidant present in the human lens; manganese concentration in the lens drops as a result of cataractogenesis[673-676]. Finally, similarly to the finding in BMES[677], there was also an association between supplemental intake of micronutrients and cross-sectional nuclear cataract but not between supplemental nutrients and cataract progression. As only 10% or less of the participants in our study took any single supplement, supplements had to be grouped together; therefore, drawing conclusions on the effect of any single supplement or of components of supplements (e.g., supplemental vitamin C) is not possible.

Most likely due to lack of power, no association was detected between any of the

---

* FFQ

† intake and plasma concentrations

micronutrients and cortical cataract progression. Vitamin D was marginally associated with increased risk of developing cortical cataract but that association would not survive a correction for multiple testing. Given the association between cortical cataract and T2D and obesity, exploring the role of macronutrients (e.g. carbohydrates) may be more fruitful in terms of associations. However, given that diet plays a much greater role in nuclear as opposed to cortical cataract formation, very larger sample sizes may be needed in order to detect any subtler effects important for cortical cataract formation.

The low power also restricted the number of covariates that could be explored. Three covariates in particular need mentioning: BMI, diabetes status and smoking. In TwinsUK, neither nuclear nor cortical cataract were cross-sectionally associated with BMI or with the presence of diabetes (p-value > 0.05). That was also true for the nutrients studied except for manganese, and, therefore, the analysis were not adjusted for BMI and diabetes status. In the case of manganese, adjusting for one or both factors did not result in a change in the beta and standard error estimates but the p-value increased (from 0.01 to 0.02). Of note, only 10% of TwinsUK participants had diabetes or were diabetes suspects, and this percentage fell to 6% within the individuals with cataract measurements. Because 85% of participants who contributed to the progression analyses have never smoked, there was also lack of power to explore the effects of smoking on cataract progression. Cross-sectionally, there was no association between smoking status and intake of micronutrients. Therefore smoking status is unlikely to have confounded the association between vitamin C and cataract.

Next, I explored whether a healthy diet, as represented by two different indexes (modified MDS, HEI) was associated with changes in blood metabolites and the microbiome. Despite the fact that the FFQ data and the metabolite/microbiome samples were obtained on average 2.4 years apart, both indices were associated with differences in blood metabolites and the microbiome. A possible reason for this is the relative stability of dietary patterns over tine; also supported by the fact that many food intake patterns and preferences are heritable[678-680]. Somewhat unexpectedly, the sets of metabolites and OTUs largely did not overlap which would suggest that the two indices reflect different aspects of the healthy diet. The MDS was better in capturing differences in metabolite concentration than HEI: higher number of metabolites both in the single point and in the mean concentration analysis were associated with MDS as opposed to HEI; the associations with MDS also reached higher significance. These results are probably due to the differences in food groups that are included in the two

scores. Conversely, more OTUs (60 vs. 20) were associated with HEI and the associations spanned all the bacterial classes. It is not clear why MDS and HEI give such different results. A varied diet is associated with high microbiome diversity[681]. The HEI measures more food groups than MDS and is probably better at measuring such diversity. HEI was also more strongly associated with alpha-diversity as opposed to MDS.

The observed associations with both scores also make biological sense. Increase in one or both scores was associated with increase in relative concentration in blood metabolites such as omega-3 fatty acids (EPA, DHA), glycerol (glycerate) and docosapentaenoate which are linked to consumption of olive/fish oil intake[682,683]; and threonate linked to consumption of nuts*. There was also a decrease in: 1,5-anhydrosorbitol concentration which is a marker of good glycemic control[684,685]; and of the carnitine metabolism by-product 3-dehydrocarnitine — a marker of reduced red meats† consumption. Of note, both scores were associated with increase in 2-aminobutyrate in the single-point and mean metabolite analysis. This metabolite is a key intermediate in the synthesis of a glutathione analog in the lens known as ophthalmic acid or ophthalmate[686]. The role that ophthalmic acid plays in the human lens is very poorly studied. The observed direction of effect, however, was not always consistent with what would be expected given a positive effect on health. For example increase in diet score was associated with increase in relative concentration of pipecolate and ketoleucine. The first metabolites is of bacterial origin and accumulates in patients with chronic liver diseases and pyridoxine-dependent seizures, while the second is a marker of compromised brain energy metabolism.

Some of metabolite associated with MDS and HEI may also be a marker of the intake of certain foods[687]: decreased relative concentration of pipecolate has been associated with intake of jam/confectionary; increased relative concentration of glycerate has been associated with tomatoes intake; increased relative concentration of EPA, DHA and DPA has been associated with various foods, including oily fish, wine, avocado; decreased relative concentration of 1,5-anhydroglucitol has been associated with intake of non-oily fish and sea food; increased relative concentrations of 2-aminobutyrate and 3-dehydrocarnitine have been associated with intake of wine and low-fat milk respectively; finally, increased relative concentration of 1-docosahexaenoylglycerophosphocholine* has been associated with all types of fish,

* http://www.hmdb.ca/metabolites

† http://www.hmdb.ca/metabolites

green leafy vegetables and avocado.

Three metabolites, pipecolate, proline and un unknown compound (X - 02269), showed association with cataract (p-value < 0.05) but none of this associations would survive correction for multiple testing. How these metabolite would affect cataract formation is unclear.

In terms of microbiome, both scores were associated with increase in total microbiome diversity with increase in both scores. HEI was associated with an increase in the Ruminococcaceae family of bacteria, which is important for digestion of plant material. The score was also associated with a decrease in Mogibacteriaceae family whose function in the gut is not well understood but some of its species are pathogens linked to gum inflammation and allergies. In this chapter weak evidence pointing to OTUs mediating the effect of diet on cataract was also presented. Two test were used, a Sobel-Goodman test and a structural equation modelling. Only the latter test detected mediation. This test is more powerful as it uses a more complex model fitting that is less susceptible to none-normality.

There may be several reasons for explaining why significant associations between metabolites/microbiome and cataract were not observed. It may be that differences in metabolites or bacterial concentrations have either no effect on the lens or the effects are sufficiently small that this study was underpowered to detect them. Another reason is that, largely only cross-sectional data was available, forcing the assumption that diet, metabolite profiles and microbiome are stable in time. In this cohort, the dietary patterns are relatively stable but if the metabolome/microbiome are not, the timing between the metabolome/microbiome sampling and the eye measures might reduce power to detect effects. Moreover, for the majority of subjects both metabolome and microbiome were sampled after the eye test. A tighter sampling time-frame might therefore be needed.

A robust association between metabolites/microbiome and cataract needs to be established before causation or mediation analysis can be tested. Finding such associations, however, will lead to better understanding of the systemic and molecular causes of cataract; and the mechanism behind the diet-cataract association. What is more, if blood metabolites mirror the metabolome of the aqueous humor, we could use blood as a model system instead of the difficult to obtain lens tissue. Changes in metabolites or the gut flora are likely to be a much faster read-out of environmental changes and thus might be used prior to disease detection or in its early stages. Therefore a relationship between those and cataract should also be studied

longitudinally. Finally, dietary interventions that target the metabolome or the microbiome may be the more powerful in preventing disease onset or progression and have other benefits rather than a cataract-specific drug therapy.

Finally, all the results presented here need replication. In addition, the TwinsUK cohort is predominantly a female cohort, and assess any gender differences in risk factors could not be assessed. The findings of this study can be generalisable only to Caucasian women of similar age because they reflect cataract progression in a group of white British women between, on average, the ages of 60 and 70 years, and so may not reflect other population groups or age ranges. The fact that a lot of individuals were lost of follow-up may have introduced a selection bias. The individuals who declined participation were of lower socioeconomic status, had higher self-rated health status, and were less health aware[464]. Any introduced bias probably would have resulted in loss of power because this group of individuals are more likely to have less healthy diets and more cataract. Those with follow-up data were on average 1.8 years younger than the original cohort, but in general they were not significantly different in other respects or in nutrient or supplement intake, hopefully reducing potential selection bias in the progression data. As in any observational study, this one is potentially susceptible to residual confounding, missing data, or misspecification of variables.

Chapter 7. General Discussion and Conclusions

7.1. Overview

The PhD thesis presented here aimed to elucidate the genetic factors and some of the environmental factors underlying the two most common forms of ARC. In this chapter the strengths and limitations of the findings presented in the previous chapters of this thesis will be summarised. The contribution this thesis makes to the existing literature will be explored and future direction of research will be considered.

7.2. General discussion

At present cataract surgery is the only form of treatment for ARC but it is not widely accessible in big parts of the world, particularly sub-Saharan Africa. In more developed economies, it is the most common operation and costs health services significant amounts of money. If a much cheaper, non-invasive treatment or a preventive strategy is to be developed, we need a better understanding of how genetic and environmental risks interplay. High-throughput 'omics' technologies can enable the discovery of the molecular changes and biological pathways that are involved in cataract formation and progression. The thesis focused on studying the genetics of ARC and a potentially modifiable risk factor — diet. The results of this thesis further the field of cataract research as follows. First, based on the genetic, environmental and dietary analyses, it can be concluded that nuclear and cortical cataract appear to be two separate diseases. Second, this thesis elucidates common genetics variants associated with nuclear cataract and proposes a list of plausible functional candidate genes that may underlay this condition. Third, the results presented in **Chapter 6** show that vitamin C intake is protective against cataract progression. Finally, the thesis attempted to address the mechanisms behind the association between diet and nuclear cataract. Below I will discuss the contribution these findings make to the literature, what implications they have for future research and care; what are the limitations; and possible future directions.

7.2.1. Age-related cataract: umbrella term for ethnologically separate conditions

A fascinating question that was addressed as part of the research presented here, and that, has never been addressed before, was whether the two common types of ARC, nuclear and cortical cataract, share genetic risk factors. In **Chapter 1**, the

different environmental risk factors that underlie nuclear and cortical cataract were discussed and it was concluded that apart from age they do not share many environmental predictors. For example, smoking and diet are associated with nuclear but not with cortical cataract, while UV radiation, T2D and being female are associated with cortical but not with nuclear cataract. The results from the diet, metabolome and microbiome analysis presented in **Chapter 6** provide further support for differences in risk profiles between the two types of cataract. However, a difference in environmental risks is not necessarily equal to difference in genetic risk. Given that both cataracts are common with advancing age, and commonly co-exist, it is widely assumed that they are related. Therefore, it was hypothesised that both common types of ARC would share at least some genetic risk factors but that the phenotypic expression will depend on environmental factors.

Here, the thinking behind this hypothesis will be explained further as it may not be immediately clear. With the exception of the embryonic nucleus, all other lens fibres are derived from the lens epithelium; and in the mature lens fibres there is no transcription but only stabilised mRNA molecules[67]. This will mean that fibres are genetically the same irrespective of their position within the lens. However, fibres in the nucleus and cortex differ in age, meaning a difference in cumulative exposures and stress-related damage, and differences in extracellular environment. Below a hypothetical example is discussed. A person is a smoker and spends a lot of time outdoors for his work. He also carries an eQTL variant that causes a decrease in concentration of a chaperon that will result in both nuclear and cortical fibres having reduced amount of the chaperon compared to non-carriers. The chaperone molecules get damaged with age and, although the same amount was present in all fibres, there are fewer molecules effective in the lens nucleus. The smoking results in additional damage and nuclear cataract but not cortical as the cortex has enough of active chaperone molecule to prevent misfolding. On the other hand, in the inferonasal quadrant area of cortex UV radiation causes damage resulting in fewer working copies of the same chaperone, misfolding and cortical cataract. In this case both cataract types share the same genetic risk but different environmental triggers.

The results of the heritability analysis presented in **Chapter 3** of this, shown that this hypothesis is incorrect. Age-related nuclear and cortical cataract are genetically distinct conditions, as shown by the results of three different approaches: a residual maximum likelihood model implemented in GCTA, linkage disequilibrium regression analysis and polygenic risk score. A strategy that could have been adopted but would have probably yielded similar results is a bivariate GWAS analysis such as the network inference in matrix-variate Gaussian models type implemented in PHENIX[688]. A

problem with this model, one that is also shared with GCTA and to a certain extent with the classical bivariate heritability estimation, is that analysis of a mixture of quantitative and binary traits results in lower power to detect association comported to two quantitative traits or two binary traits. Cortical cataract is not easily transformable due to a high number of unaffected individuals (score of 0) as using rank transformation which usually works for log tailed data results in people with scores of zero getting a positive ranking.

The question is then what makes nuclear and cortical cataract genetically different. One possibility is that very different sets of genes are active in young and old fibres but studies of molecular differences between the human lens nucleus and the cortex are lacking, probably due to the difficulty of physically separating the two. A more interesting possibility is that somatic mutations accumulate in the lens epithelium as the eye ages and are passed to the fibres[689]. Whatever the reason, an implication is that the same prevention strategy or a future non-operative treatment is unlikely to work for both. Further, another research question arises, namely, if the two common types of cataract do not share genetic risk factors, do each of them share risk factors with other common-age related diseases? For example, does cortical cataract shares genetic risk factors with diabetes and metabolic syndrome; or does nuclear cataract share genetic risk factors with telomere length?

These findings would benefit from a larger sample size for a more accurate estimation and also from a replication in an independent cohort. However, given that there are almost no variants, loci or genes overlapping between the two types of age related cataract, neither in the TwinsUK, nor between the age-related nuclear cataract meta-analysis and similar meta-analyses of genetic studies of age-related cortical cataract, the results are not expected to change substantially. Finally, this also means that cataract surgery cannot be used as a surrogate for either nuclear or cortical cataract diagnoses. Neither can self-reported status of either having the disease or having had cataract surgery be used. For example, in TwinsUK, about 50% of volunteers who responded that they did not have cataract actually were affected with “clinically significant” nuclear or cortical cataract, or with both types of cataract. Unfortunately, the large population-base studies such as UK Biobank, do not have information on cataract types or cataract grading. Due to lack of time in busy clinics, such information is often missing from the clinical record too. That also makes it extremely difficult to obtain large sample sizes unless a cataract-specific study is designed to do that. All that said, it might be possible to exploit the differences in risk (both genetic and environmental) between the two conditions two differentiate between nuclear and cortical cataract in patients diagnoses of cataract where information on

cataract type is not available. How that can be done is discussed in the **Future Studies** section below.

7.2.2. Genetic risk factors for nuclear cataract

Age-related cataracts are common complex diseases and as such arise from the interplay of genetic and environmental factors. In the case of ARC, environmental risk factors have been studied extensively, as summarised in **Chapter 1**, but apart of couple of heritability studies[167-170], little focus has been given to genetic risk factors[122,574]. Thus, in **Chapters 4** and **5**, it was hypothesised that there exist genetic variants (common and rare in terms of effect allele frequency) that are associated with the two common forms of ARC, nuclear and cortical cataract. Through my efforts and those of my colleagues of the International Cataract Genetics Consortium, new susceptibility loci for nuclear and cortical cataract have been identified. As my efforts were primarily devoted on elucidating the genetic risk factors for nuclear, rather then cortical cataract, the following discussion will be focused this on nuclear cataract.

In terms of nuclear cataract genetics, main contribution to knowledge this thesis makes, is the identification of 6 new loci (*SOX2-OT, TMPRSS5, COMMD1, LINC01412, MMAB, GLTSCR1*) associated with age-related nuclear cataract and the replication of the previously reported association with variants in the *CRYAA* gene[574] (**Chapter 4**). Apart from *MIP* ($P=2.4x10^{-4}$), none of the ARC candidate genes put forwards by various small studies[64,178-205] were found to be associated with nuclear cataract (p-value < 0.05). These results again prove that many studies adopting a candidate gene strategy are poorly powered to detect true associations and there is likely a positive publication bias. In this thesis it was also shown for the first time that, unlike most age-related eye-phenotypes, there is a genetic link between age-related and congenital forms. Coding mutations in genes such as *CRYAA*, *CRYAB*, *SOX2* and *GJA3* cause congenital cataracts[59,60,588,592], while, similarly to other age-related diseases[690,691], variants associated with age-related nuclear cataract (**Chapter 5**) are more likely to affect gene expression. It is likely that while mutations in crystallins or other lens proteins cause rapid and direct protein aggregation and result in congenital cataract, common polymorphisms may increase or decrease gene expression and thus susceptibility to environmental insults, such as oxidative stress, and result in ARC by altering the pool of available molecules[28,126,692]. The functional work (RNA sequencing and protein expression) carried by our collaborators had only a small set of human lens tissues (predominantly epithelial cells) available and would benefit from a further replication. Nonetheless it provided supportive evidence for cataract-related differences in RNA

and protein expression for genes at the identified loci. Given that lens fibres arise from the epithelial cells and that mature fibres are transcriptionally incompetent and relatively inactive metabolically[28], altered expression patterns of key genes in lens epithelial cells are likely to propagate to the mature lens fibres. Although nuclear cataract has onset in the 5th decade of life and nuclear scatter detectable prior early adulthood, the lens nucleus is already present at birth and, therefore, subtle dysregulation of the development process described in **Chapter 1** (Section 1.4. Anatomy and embryonic development of the lens) may be of very important but might not be detectable adult lens epithelium.

Based on the bioinformatics work presented in **Chapter 5**, together with evidence from the existing literature, plausible candidate gene list for nuclear cataract can now be drawn. The lens gene signature[693] (relative to other eye tissues) will be used as a starting point to explain a possible role of these candidate genes in nuclear cataract formation. The idea being that if a gene is strongly (or exclusively) expressed in lens it should be important for its function. In this discussion genes will also be include that, due to lack of power, did no reached suggestive or genome-wide significance but have strong a priori probability to be important for cataract formation and showed (in the meta-analysis) statistical evidence of association of varying strength. Due to the number of multiple tests performed, strict Bonferroni correction was applied which is generally very conservative if the independence assumption does not hold, potentially overlooking true associations.

Within the associated loci there were genes (*CRYAA*, *CRYAB,* and genes within the *CRYB* and *CRYG* clusters) that fall within the structural lens proteins. It is easy to see how genetic variants that cause differences in expression levels or protein instability, or in the case of free form of CRYAA and CRYAB, a decrease in chaperone activity, may cause protein unfolding and lens clouding[370,586,587,694]. Regarding the other structural protein encoding genes (*LIM2*, *BFSP2*, *SPTBN2*, *ABLIM, EPB49, EPB41L1, EPB41L4*), enriched in the lens signature, variants in close proximity to *LIM2* and *BFSP2* were also found to be associated here. *LIM2* encodes for a protein with a key role in cell-cell adhesion, while *BFSP2* encodes for a protein that, along with filensin, composes the beaded filaments – lens cell-specific intermediate filaments[695,696]. Although not signature proteins, but still very important for lens function[64,584] variants altering the function of genes (*GJA3* and *GJA8*) encoding for gap junction proteins also may be involved in formation of age-related nuclear cataract. Another notable feature of the lens signature is the enrichment of genes encoding proteins involved in endocytosis, including clathrin (*CLTCL1*, *PICALM*) and caveolin (*CAV1*). None of those genes were found associated with nuclear cataract. Of note, of the genes in the

GLTSCR1 locus, *NAPA* expression was associated with cataract severity. *NAPA* encodes for a protein of the SNAP family of vesicle docking proteins[697].

Given that two (*SOX2*, *ZEB2*) of the candidate genes found to be associated with nuclear cataract are transcription factors, regulating the transcription of lens genes (structural and others), transcriptional regulation may also play an important role in the formation of nuclear cataract[593-595,597,698]. Furthermore, the current study suggest that the gene encoded by *EHBP1* is Eps15 homology domain binding protein that interacts with EHD1 – a protein important for regulation of lens development[603,604], may also play a role in nuclear cataract pathogenesis. The involvement of transcription factors may mean that prenatal and early-life exposure might affect the way the lens ages. Alternatively, transcription factors may be re-expressed in the lens epithelia which in turn might affect the lens fibres derived from the epithelia. Such is the is the case with SOX2 which is activated in a specific population of adult lens epithelial cells and is required for their self-renewal[699].

The majority of the previous candidate gene studies focused on genes important for protection against oxidative stress in line with the oxidative stress hypothesis of cataract formation[126]. Several classes of such genes are also within the lens expression signature set: heath-shock proteins (*HSPA6*, *HSPA8*, *HSPB1*), proteasome complex subunits (*PSMA6*, *PSMA7*, *PSMB6*, *PSMB7*, *PSMB9*, *PSMD13*), and scavenging of free radicals (*GSS*, *GSR*, *SOD1*, *AOP2*)[693]. The oxidative stress hypothesis was formed based on functional studies in animal models and cell culture experiments; and more rarely human lens samples. An additional compelling piece of evidence however is the fact that patients develop nuclear cataracts secondary to treatment with hyperbaric oxygen[700]. According to this hypothesis, ROS are formed in the lens either endogenously (as result of the function of mitochondria (lens epithelium), peroxisomes, lipooxygenases, NADPH oxidase and cytochrome P450, lens specific forms of NADPH oxidase located in the plasma membrane of lens fibres) or exogenously (sources include: ultraviolet light, ionising radiation, chemotherapeutics, inflammatory cytokines and environmental toxins)[157]. When the lens anti-oxidation mechanisms are overwhelmed or do not function properly, ROS can damage various lens fibre components. For example ROS oxidise lens proteins and makes them insoluble or impair membrane stability by oxidising membrane lipids[157,700-702]. ROS can also increase ion balance in the lens epithelium if it damages the Na/K-ATPase pumps[703]. When metals such as Fe and Cu that have been found to be elevated in advanced cataract react with H_2O_2 that also can lead to the production of extremely potent ROS[700]. Apart from variants in close proximity to *SOD1* (p-value = 5.4×10^{-4}), none of the above listed genes that when mutated have been proposed to decrease

the lens ability to defend agains ROS, were found to be associated in the meta-analysis of GWAS of nuclear cataract. Although we did not find association with those genes, variants in these and other genes with anti-oxidative function may still play some role and may be detected with a better-powered study.

All these results, and the fact that the associated genes in general are not known risk factors for other systemic conditions, point to more eye-specific genetic causes for nuclear cataract as opposed to genes associated with general body ageing and metabolic imbalance. As with the genes linked to oxidative stress, this observation may be due to insufficient power to detect effects.

Although the International Cataract Genetics consortium has the largest to date sample size and some of the best phenotypically defined studies, it lacked the power to elucidate additional loci which will lead to better understanding of both types of cataract. In addition, reducing the heterogeneity of grading system by re-grading studies and using qualitative (e.g. central nuclear dip score[167], dynamic light scattering[694]) rather than semi-qualitative grading systems may afford additional power.

Finally, more functional work is needed to point with certainty the causative genes, the possible regulatory or other effects mediated by the variant (or variants in LD), in each of the loci identified in this thesis. Even though functional work in the human lens is made difficult by the fact that modern cataract surgery operations do not provide intact lens tissues, it still is likely to add to the usefulness of animal models. Any insight from model systems will need validation in human tissue. Pioneering work is exploring the possibility of growing a lens from an individual's own stem cells, including lens epithelium stem cells[704,705]. If those prove to be close in biochemical priorities to the person's own lens, lenses derived from pluripotent stem cells might prove to be a valuable and easier alternative to post-mortem lens material.

7.2.3. Diets and cataract

The focus of my research prior to pursuing a PhD degree has always been primarily in different aspects of human genetics (including molecular genetics, clinical genetics, genetic epidemiology). Thus, my first year rotation project was in the field of evolutionary genetics and dealt with the question of human adaptation to Neolithic dietary change. Apart of a valuable experience in a field of genetics I have never before worked in, this project sparked in me a fascination with the role diet plays in health and disease. In my PhD project I combined this fascination with eye research. There is plenty of evidence (discussed in detail in **Chapter 1**) in the literature showing that high intake of antioxidant micronutrients protects against cataract formation. Whether this is

also true about cataract progression is not certain. A small subset of the TwinsUK cohort had both baseline and follow-up cataract measurements, and those could be used this data to answer two questions: 1) what is the heritability of cataract progression; 2) does micronutrient intake protect against cataract progression. I showed in this thesis that genetic factors explained 35% of the variation in 10-year progression of nuclear cataract that dietary vitamin C intake at baseline protected against 10-year cataract progression. The meta-analysis of GWAS of nuclear cataract is probably the best achievement (in scientific terms) presented in this thesis but the vitamin C work is the one with the biggest public impact. The article[656] presenting the results of the vitamin C work was extensively covered in the British and international media, including in the New York Times. The reason for such impact is that a simple intervention, in this case increasing the intake vitamin C, can protect not only against cataract formation but also against cataract progression.

Given that the different diet components are correlated and that dietary patterns or preferences[706], as opposed to single nutrients, may reflect better the effect of diets as a whole, next the following questions were explored: whether "healthy" diet has an effect on nuclear cataract and whether the effect is mediated by metabolites or the microbiome. The rationale behind exploring those questions was that if diet or dietary factors protected against cataract formation and or progression, and that effect was mediated by changes of metabolites or the microbiome, then it could be explored whether any of the genetic risk factors for ARC interact either with diet (dietary factors) directly or indirectly through the metabolome or the microbiome. Assuming that such interaction exists and can be detected in blood, one could devise a treatment that modify the effect of one or more cataract-related genes by modifying a person's diet and, thus, to stop formation or slow progression of ARC.

The results showed that "healthy diet" as measured by two different indices was protective against cataract formation. Moreover, diet was also associated with changes in metabolites and the microbiome. Unfortunately, there was not a strong evidence of association between diet-related metabolites/OTUs and age-related nuclear cataract. There may be a number of different reasons for that. First, it might be that due to the blood-eye barrier the metabolic environment of the eye is very different from that in blood and so blood is the wrong tissue to measure. However, as long as the concentration gradient is not reversed or absent an effect should still be detectable. For example, the concentration of ascorbic acid is substantially higher in the aqueous humor than in blood but an association between plasma ascorbic acid and cataract is detectable. Probably a stronger factor is the gap between the metabolome/microbiome sampling and the eye test where the latter was done up to 5 years after the former.

Although dietary patterns, metabolites and the gut microbiome seem stable in the long-term[707] and cataract is not reversible, other factors that affect both cataract and metabolites/microbiome might confound a possible association. For example during the eye visit the participant would have been made aware of their cataract status and that could have resulted in change of exposure patterns (e.g. stopped smoking) that were not taken into account. The change could have caused concentration differences in metabolites but not in cataract status which will prevent an association being detected. A tighter matching of sampling dates would afford better power to detect association. Further to that, using platforms measuring more metabolites or doing so with increased accuracy (eg NMR-based platforms) will likely generate additional findings and will make it possible to explore the links between diet either as a whole per nutrient (e.g. vitamins), metabolites and cataract. Although the lack of statistically significant associations between metabolites/microbiome and ARC is disappointing, further studies are worthwhile because if a pathway linking diet and cataract through metabolites is established this could pave the way for new interventions. If diet/dietary patterns as a whole have more influence that separate nutrients, it might also explain why randomised controlled trials for supplemental vitamin intake largely failed. A different approach can also be taken, for example by exploring whether genetic variants linked to cataract are also associated with metabolites or OTUs. Alternatively, multivariate approaches such as building networks of diets, metabolite and OTUs and associating those with cataract may prove more fruitful albeit much less straight forward to interpret.

Finally, the results of the dietary epidemiology part of this thesis need replication in an independent dataset and a bigger sample size would afford power to study the role diet in cataract progression. In addition, the TwinsUK cohort is predominantly a British female cohort which reduces generalisation to other population groups or age ranges. One of the most important limitations of the analysis presented in **Chapter 6** is the fact that they dealt with associations which is very different from causations. Associating one latent variable with another does not imply causation or its direction. Therefore, in the case of the association between diet, microbiome and cataract, mediation analyses were attempted but due to lack of power the results were inconclusive. A larger study would afford the power to perform not only mediation analysis but also Mendelian randomisation. And of course, a randomised clinical trial would be in a better position to elucidate causality. The TwinsUK is a representative population cohort but as any observational cohort, the results presented here are potentially susceptible to residual confounding, missing data, or misspecification of variables. This is especially true for the progression part of the study.

7.3. Future directions

As discussed in this chapter, nuclear and cortical cataract seem to be driven by distinct combinations of genetic and environmental risk factors while in majority of cases clinical records contain only a diagnose of cataract without its type being specified. Similarly, records on cataract surgery that do not mention the type of cataract observed before the surgery are also of little use. These ambiguities seriously limit the power of any cataract related studies. However, cataract is not the only field suffering from problems relating to discrimination of cases. For example, close to half of all type 1 diabetes (T1D) cases have onset over 30 years of age that overlaps with the onset of T2D. Those two groups of patients are difficult to distinguish clinically. Recent studies however, have shown the utility of genetic risk score for distinguishing between adult onset T1D and T2D[708,709]. A similar technique could be used in the case of nuclear and cortical cataract as described below.

As part of the UK Biobank imaging study (N = 100,000), lens photos could be obtained and both types of cataract graded using, ideally, objective grading systems. As these research participants are already genotyped, a genetic risk score can be derived either from well measured common SNPs or from SNPs that are conditionally independently associated with each type of cataract. The effects of rare variants from the sequencing data that would soon be available for all UK Biobank participants could also be added. The score can be augmented by data on environmental risks. A score that discriminate well between nuclear and cortical cataract cases using a reasonably small set of genetic variants and is replicated in large independent cohorts such as the cohorts from the International Cataract Genetics Consortium, can open the doors to using almost any population or hospital-based study irrespective of how cataract presence was recorded. That would includ studies such as the Kayser Permanente GERA study in California or the Million veterans program, that use the advent of huge electronic medical record studies linked to DNA. It will also make it possible to use studies that do not have any eye data at all but have a wealth of molecular phenotyping data available such as INTERVAL (www.intervalstudy.org.uk/). In such cases the score could predict whether a person is likely to develop cataract of each type and controls can be individuals who scores low on both cataract scores. In terms of ethnic differences, the trans-ethnic meta-analysis of age-related nuclear cataract presented in this thesis do not support differences in genetic risk between European and Asian ancestry cohorts and thus a genetic risk score might be generalisable to other ethnicities.

All this will make it possible to study the genetics of ARC in large cohorts and

perform a much more comprehensive and hugely-powered analysis of genetic variation through the entire frequency spectrum including also CNVs or poorly explored parts of the genome such as the X chromosome and mitochondrial DNA variants. Other various types of analysis that require big sample sizes will be possible such as gene by gene or gene by environment interaction studies. Studying gene-by-environment and gene-by-gene interactions requires huge sample sizes that are prohibitive for hypothesis-free approaches and require prior knowledge about underlying genetic and environmental risk factors[710-712].

When more loci can be identified and causative genes precisely fine-mapped in conjunction with metabolomics and proteomic data that would afford deeper knowledge of molecular mechanisms underlying ARC and hopefully point to better preventive and therapeutic strategies.

7.4. Potential cataract therapies

Potential therapeutic and preventive compounds for ARC are actively explored but there have been two main problems. First, the limited knowledge on the mechanisms underlying the disease and the lack of genetic studies has made it difficult to guid drug targeting or re-purposing. Thus there are just a few examples of therapeutic agents that show some promise. In recent studies, triterpenoids (lanosterol and a related compound named compound 29) improved lens transparency by reducing aggregation of crystalline proteins[713,714]. A later study however showed that lanosterol did not reverse opacification in patients with age-related nuclear cataract[715]. More recently, ghrelin, the so called hunger hormone, has also been shown to protect rat lens epithelial cells agains oxidative stress[716].

Lanosterol was identified in a congenital cataract genetic study thus showing the promise of using genetics tools for drug discovery[714]. Other fields, where association studies in very large sample have been possible have had much more success in inferring drug properties from genetic association results in combination with Mendelian randomisation approaches. A good examples is darapladib, an inhibitor of a LDL-bound enzyme expressed by inflammatory cells in atherosclerotic plaques, and thus a promising anti-coronary heart disease drug[717]. Darapladib failed tree clinical trials but it was only later that a large-scale human genetic study found that none of a series of enzyme-lowering alleles was causally related to coronary heart disease risk[717].

It remains to be seen however whether any new cataract therapeutic approach, unless derived from repurposing, can reduce cataract-related visual impairment globally, given that new agents may have limited application in developing countries

due to drug pricing and access issues, which may place them out of reach of those who are most in need. In the developing world preventive treatments may, therefore, be more successful. In the past decade an enormous effort has been made to elucidate the anti-oxidant properties of plant extracts in experimental systems with the hope to create preventive, possibly dietary, interventions for cataract[718]. Compounds extracted form vast variety of plants, from every-day foods (apricots, garlic, curcuma) to exotic plants (silk-cotton tree, drumstick tree, Chinese chaste tree) have been tested[719-724]. But it is still unclear how effective those can be when administered to humans and in majority of cases the precise mechanisms underpinning the observed protective effects are unknown. As with a lot of other nutrient-based interventions, there is not an easy read out available in terms of the target or systemic effects once the compound is ingested. A better understanding of the human blood metabolome and gut microbiome and their connection to the eye, however, may possibly provide such read outs and such studies are thus worth perusing.

7.5. Conclusion

In conclusion this PhD thesis elucidated some of the genetic factors underlying ARC and especially age-related nuclear cataract. It also attempted to find the mechanisms behind the association between diet and cataract and showed that vitamin C protects against cataract progression. The work presented in this thesis together with the other projects pursued during my postdoctoral studies is not the end of a road but a stepping stone for future research. Cataract is the leading cause of world blindness but the complex interplay between the multitude of common genetic variants that are likely to confer cataract risk and environmental factors is still a *terra incognita.* In order to study this interplay we will need better powered purpose-designed studies as well as biomarkers that could help us to distinguish molecularly between the common types of ARC.

References

1. Pascolini, D. & Mariotti, S.P. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology* **96**, 614-618 %U http://bjo.bmj.com/cgi/doi/10.1136/bjophthalmol-2011-300539 (2012).

2. Hong, T. *et al.* Visual impairment and depressive symptoms in an older Australian cohort: longitudinal findings from the Blue Mountains Eye Study. *Br J Ophthalmol* **99**, 1017-21 (2015).

3. Bourne, R.R. *et al.* Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990-2010. *Br J Ophthalmol* **98**, 629-38 (2014).

4. Keeffe, J. *et al.* Prevalence and causes of vision loss in Southeast Asia and Oceania: 1990-2010. *Br J Ophthalmol* **98**, 586-91 (2014).

5. Khairallah, M. *et al.* Number of People Blind or Visually Impaired by Cataract Worldwide and in World Regions, 1990 to 2010. *Invest Ophthalmol Vis Sci* **56**, 6762-9 (2015).

6. Kuper, H., Polack, S. & Limburg, H. Rapid assessment of avoidable blindness. *Community Eye Health* **19**, 68-9 (2006).

7. Lee, C.M. & Afshari, N.A. The global state of cataract blindness. *Curr Opin Ophthalmol* **28**, 98-103 (2017).

8. Hansen, M.S. & Hardten, D.R. Financially efficient cataract surgery in today's healthcare environment. *Curr Opin Ophthalmol* **26**, 61-5 (2015).

9. Minassian D. C. , R.A. Future sight loss UK (2): an epidemiological and economic model for sight loss in the decade 2010–2020. **2009**, 1–130 (2009).

10. Ibrahim, N., Pozo-Martin, F. & Gilbert, C. Direct non-medical costs double the total direct costs to patients undergoing cataract surgery in Zamfara state, Northern Nigeria: a case series. *BMC Health Serv Res* **15**, 163 (2015).

11. Al Ghamdi, A.H. *et al.* Rapid assessment of avoidable blindness and diabetic retinopathy in Taif, Saudi Arabia. *Br J Ophthalmol* **96**, 1168-72 (2012).

12. Bettadapura, G.S. *et al.* Assessment of avoidable blindness using the rapid assessment of avoidable blindness methodology. *N Am J Med Sci* **4**, 389-93 (2012).

13. Kandeke, L. *et al.* Rapid assessment of avoidable blindness in two northern provinces of Burundi without eye services. *Ophthalmic Epidemiol* **19**, 211-5 (2012).

14. Lindfield, R., Griffiths, U., Bozzani, F., Mumba, M. & Munsanje, J. A rapid assessment of avoidable blindness in Southern Zambia. *PLoS One* **7**, e38483 (2012).

15. Marmamula, S., Narsaiah, S., Shekhar, K. & Khanna, R.C. Visual impairment among weaving communities in Prakasam district in South India. *PLoS One* **8**, e55924 (2013).

16. Patil, S. *et al.* Prevalence, causes of blindness, visual impairment and cataract

surgical services in Sindhudurg district on the western coastal strip of India. *Indian J Ophthalmol* **62**, 240-5 (2014).

17. Randrianaivo, J.B. *et al.* Blindness and cataract surgical services in Atsinanana region, Madagascar. *Middle East Afr J Ophthalmol* **21**, 153-7 (2014).

18. Hajar, S., Al Hazmi, A., Wasli, M., Mousa, A. & Rabiu, M. Prevalence and causes of blindness and diabetic retinopathy in Southern Saudi Arabia. *Saudi Med J* **36**, 449-55 (2015).

19. Morchen, M. *et al.* Prevalence of blindness and cataract surgical outcomes in Takeo Province, Cambodia. *Asia Pac J Ophthalmol (Phila)* **4**, 25-31 (2015).

20. Zatic, T. *et al.* Rapid assessment of avoidable blindness and diabetic retinopathy in Republic of Moldova. *Br J Ophthalmol* **99**, 832-6 (2015).

21. Marmamula, S., Khanna, R.C., Kunkunu, E. & Rao, G.N. Population-based assessment of prevalence and causes of visual impairment in the state of Telangana, India: a cross-sectional study using the Rapid Assessment of Visual Impairment (RAVI) methodology. *BMJ Open* **6**, e012617 (2016).

22. Thoufeeq, U. *et al.* First Rapid Assessment of Avoidable Blindness Survey in the Maldives: Prevalence and Causes of Blindness and Cataract Surgery. *Asia Pac J Ophthalmol (Phila)* (2017).

23. Zhang, X. *et al.* Prevalence of visual impairment and outcomes of cataract surgery in Chaonan, South China. *PLoS One* **12**, e0180769 (2017).

24. Szabo, D. *et al.* Visual impairment and blindness in Hungary. *Acta Ophthalmol* **96**, 168-173 (2018).

25. West, S.K. & Valmadrid, C.T. Epidemiology of risk factors for age-related cataract. *Surv Ophthalmol* **39**, 323-34 (1995).

26. Shiels, A. *et al.* The EPHA2 gene is associated with cataracts linked to chromosome 1p. *Mol Vis* **14**, 2042-55 (2008).

27. Liao, J. *et al.* Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract. *Human Molecular Genetics* **23**, 6119-6128 %U http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddu315 (2014).

28. Michael, R. & Bron, A.J. The ageing lens and cataract: a model of normal and pathological ageing. *Philos Trans R Soc Lond B Biol Sci* **366**, 1278-92 (2011).

29. Spalton, D.J., Hitchings, R. A., Hunter, P. A. Atlas of clinical ophthalmology, 3rd edition. *Archives of Ophthalmology* **124**, 925-925 (2006).

30. Asbell, P.A. *et al.* Age-related cataract. *Lancet* **365**, 599-609 (2005).

31. Vrensen, G. & Willekens, B. Biomicroscopy and scanning electron microscopy of early opacities in the aging human lens. *Invest Ophthalmol Vis Sci* **31**, 1582-91

(1990).

32. Cotlier, E. Cataract Image Analysis System. *Eye (Lond)* **13 (Pt 3b)**, 457-63 (1999).

33. Chew, E.Y. *et al.* Evaluation of the age-related eye disease study clinical lens grading system AREDS report No. 31. *Ophthalmology* **117**, 2112-9 e3 (2010).

34. Sasaki, K. *et al.* Classification System for Cataracts. *Ophthalmic Research* **22**, 46-50 (1990).

35. Chylack, L.T., Jr. *et al.* Lens opacities classification system II (LOCS II). *Arch Ophthalmol* **107**, 991-7 (1989).

36. Chylack, L.T., Jr. *et al.* The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol* **111**, 831-6 (1993).

37. Sparrow, J., Bron, A., Brown, N., Ayliffe, W. & Hill, A. The Oxford Clinical Cataract Classification and Grading System. *Int Ophthalmol* **9**, 207-225 (1986).

38. Taylor, H.R. & West, S.K. The clinical grading of lens opacities. *Aust N Z J Ophthalmol* **17**, 81-6 (1989).

39. Klein, B.E.K., Magli, Y., Neider, M.W. & Klein, R. Wisconsin System for Classification of Cataracts from Photographs, (University of Wisconsin, 1989).

40. Hall, A.B., Thompson, J.R., Deane, J.S. & Rosenthal, A.R. LOCS III versus the Oxford Clinical Cataract Classification and Grading System for the assessment of nuclear, cortical and posterior subcapsular cataract. *Ophthalmic Epidemiol* **4**, 179-94 (1997).

41. Hall, N.F., Lempert, P., Shier, R.P., Zakir, R. & Phillips, D. Grading nuclear cataract: reproducibility and validity of a new method. *Br J Ophthalmol* **83**, 1159-63 (1999).

42. Tan, A.C. *et al.* Cataract prevalence varies substantially with assessment systems: comparison of clinical and photographic grading in a population-based study. *Ophthalmic Epidemiol* **18**, 164-70 (2011).

43. Pescosolido, N., Barbato, A., Giannotti, R., Komaiha, C. & Lenarduzzi, F. Age-related changes in the kinetics of human lenses: prevention of the cataract. *Int J Ophthalmol* **9**, 1506-1517 (2016).

44. Danysh, B.P. & Duncan, M.K. The lens capsule. *Exp Eye Res* **88**, 151-64 (2009).

45. Bassnett, S., Shi, Y. & Vrensen, G.F. Biological glass: structural determinants of eye lens transparency. *Philos Trans R Soc Lond B Biol Sci* **366**, 1250-64 (2011).

46. Augusteyn, R.C. Growth of the human eye lens. *Mol Vis* **13**, 252-7 (2007).

47. Xu, J., Pokorny, J. & Smith, V.C. Optical density of the human lens. *J Opt Soc Am A Opt Image Sci Vis* **14**, 953-60 (1997).

48. Cheng, R., Lin, B. & Ortwerth, B.J. Separation of the yellow chromophores in individual brunescent cataracts. *Exp Eye Res* **77**, 313-25 (2003).

49. Jahngen-Hodge, J., Cyr, D., Laxman, E. & Taylor, A. Ubiquitin and ubiquitin conjugates in human lens. *Exp Eye Res* **55**, 897-902 (1992).

50. Borchman, D., Byrdwell, W.C. & Yappert, M.C. Regional and age-dependent differences in the phospholipid composition of human lens membranes. *Invest Ophthalmol Vis Sci* **35**, 3938-42 (1994).

51. Borchman, D., Yappert, M.C. & Afzal, M. Lens lipids and maximum lifespan. *Exp Eye Res* **79**, 761-8 (2004).

52. Huang, L. *et al.* Human lens phospholipid changes with age and cataract. *Invest Ophthalmol Vis Sci* **46**, 1682-9 (2005).

53. Whitman, J.K., Alviar, A.F., Fleschner, C.R. & Stuart, M.K. Monoclonal antibody 10A5 recognizes an antigen unique to the water-insoluble 25/45 membrane fraction of the rat ocular lens. *Springerplus* **2**, 500 (2013).

54. Harding, J.J. Viewing molecular mechanisms of ageing through a lens. *Ageing Res Rev* **1**, 465-79 (2002).

55. Augusteyn, R.C. On the growth and internal structure of the human lens. *Exp Eye Res* **90**, 643-54 (2010).

56. Shanske, A.L., Bogdanow, A., Shprintzen, R.J. & Marion, R.W. The Marshall syndrome: report of a new family and review of the literature. *Am J Med Genet* **70**, 52-7 (1997).

57. Litt, M. *et al.* Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Human Molecular Genetics* **7**, 471-474 (1998).

58. Berry, V., Francis, P., Kaushal, S., Moore, A. & Bhattacharya, S. Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nat Genet* **25**, 15-7 (2000).

59. Berry, V. *et al.* Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. *American Journal of Human Genetics* **69**, 1141-1145 (2001).

60. Li, Y., Wang, J., Dong, B. & Man, H. A novel connexin46 (GJA3) mutation in autosomal dominant congenital nuclear pulverulent cataract. *Molecular Vision* **10**, 668-671 (2004).

61. Van Agtmael, T. *et al.* Dominant mutations of Col4a1 result in basement membrane defects which lead to anterior segment dysgenesis and glomerulopathy. *Hum Mol Genet* **14**, 3161-8 (2005).

62. Ruiz-Ederra, J. & Verkman, A.S. Accelerated cataract formation and reduced

lens epithelial water permeability in aquaporin-1-deficient mice. *Invest Ophthalmol Vis Sci* **47**, 3960-7 (2006).

63. Hansen, L. *et al.* Genetic heterogeneity in microcornea-cataract: five novel mutations in CRYAA, CRYGD, and GJA8. *Invest Ophthalmol Vis Sci* **48**, 3937-44 (2007).

64. Zhou, Z. *et al.* Genetic variations in GJA3, GJA8, LIM2, and age-related cataract in the Chinese population: a mutation screening study. *Mol Vis* **17**, 621-6 (2011).

65. Lovicu, F.J. & Robinson, M.L. *Development of the ocular lens*, (Cambridge University Press, Cambridge, 2004).

66. Cvekl, A., McGreal, R. & Liu, W. Lens Development and Crystallin Gene Expression. in *Progress in Molecular Biology and Translational Science*, Vol. 134 129-167 %@ 978-0-12-801059-4 %U http://linkinghub.elsevier.com/retrieve/pii/S1877117315000721 (Elsevier, 2015).

67. Cvekl, A. & Ashery-Padan, R. The cellular and molecular mechanisms of vertebrate lens development. *Development* **141**, 4432-47 (2014).

68. Patthey, C. & Gunhaga, L. Signaling pathways regulating ectodermal cell fate choices. *Exp Cell Res* **321**, 11-6 (2014).

69. Wolf, L. *et al.* Histone posttranslational modifications and cell fate determination: lens induction requires the lysine acetyltransferases CBP and p300. *Nucleic Acids Res* **41**, 10199-214 (2013).

70. Wawersik, S. *et al.* BMP7 acts in murine lens placode development. *Dev Biol* **207**, 176-88 (1999).

71. Sjodal, M., Edlund, T. & Gunhaga, L. Time of exposure to BMP signals plays a key role in the specification of the olfactory and lens placodes ex vivo. *Dev Cell* **13**, 141-9 (2007).

72. Beebe, D. *et al.* Contributions by members of the TGFbeta superfamily to lens development. *Int J Dev Biol* **48**, 845-56 (2004).

73. Rajagopal, R. *et al.* The type I BMP receptors, Bmpr1a and Acvr1, activate multiple signaling pathways to regulate lens formation. *Dev Biol* **335**, 305-16 (2009).

74. Greiling, T.M. & Clark, J.I. New insights into the mechanism of lens development using zebra fish. *Int Rev Cell Mol Biol* **296**, 1-61 (2012).

75. Pontoriero, G.F. *et al.* Cell autonomous roles for AP-2α in lens vesicle separation and maintenance of the lens epithelial cell phenotype. *Developmental Dynamics* **237**, 602-617 (2008).

76. Wurm, A., Sock, E., Fuchshofer, R., Wegner, M. & Tamm, E.R. Anterior segment dysgenesis in the eyes of mice deficient for the high-mobility-group transcription factor

Sox11. *Exp Eye Res* **86**, 895-907 (2008).

77. Cvekl, A. & Callaerts, P. PAX6: 25th anniversary and more to learn. *Exp Eye Res* **156**, 10-21 (2017).

78. Machon, O. *et al.* Lens morphogenesis is dependent on Pax6-mediated inhibition of the canonical Wnt/beta-catenin signaling in the lens surface ectoderm. *Genesis* **48**, 86-95 (2010).

79. Anand, D. & Lachke, S.A. Systems biology of lens development: A paradigm for disease gene discovery in the eye. *Exp Eye Res* **156**, 22-33 (2017).

80. Hogan, B.L. *et al.* Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. *J Embryol Exp Morphol* **97**, 95-110 (1986).

81. van Raamsdonk, C.D. & Tilghman, S.M. Dosage requirement and allelic expression of PAX6 during lens placode formation. *Development* **127**, 5439-48 (2000).

82. Collinson, J.M. *et al.* Primary defects in the lens underlie complex anterior segment abnormalities of the Pax6 heterozygous eye. *Proc Natl Acad Sci U S A* **98**, 9688-93 (2001).

83. Glaser, T. *et al.* PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet* **7**, 463-71 (1994).

84. van Heyningen, V. & Williamson, K.A. PAX6 in sensory development. *Hum Mol Genet* **11**, 1161-7 (2002).

85. Oliver, G. *et al.* Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* **121**, 4045-55 (1995).

86. Bovolenta, P., Mallamaci, A., Puelles, L. & Boncinelli, E. Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors. *Mech Dev* **70**, 201-3 (1998).

87. Ashery-Padan, R., Marquardt, T., Zhou, X. & Gruss, P. Pax6 activity in the lens primordium is required for lens formation and for correct placement of a single retina in the eye. *Genes Dev* **14**, 2701-11 (2000).

88. Lang, R.A. Pathways regulating lens induction in the mouse. *Int J Dev Biol* **48**, 783-91 (2004).

89. Kamachi, Y. *et al.* Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J* **14**, 3510-9 (1995).

90. Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R. & Kondoh, H. Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development* **125**, 2521-32 (1998).

91. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet* **16**, 182-7 (2000).

92. Kondoh, H., Uchikawa, M. & Kamachi, Y. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *Int J Dev Biol* **48**, 819-27 (2004).

93. Furuta, Y. & Hogan, B.L. BMP4 is essential for lens induction in the mouse embryo. *Genes Dev* **12**, 3764-75 (1998).

94. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev* **15**, 1272-86 (2001).

95. Donner, A.L., Episkopou, V. & Maas, R.L. Sox2 and Pou2f1 interact to control lens and olfactory placode development. *Dev Biol* **303**, 784-99 (2007).

96. Tantin, D., Schild-Poulter, C., Wang, V., Hache, R.J. & Sharp, P.A. The octamer binding transcription factor Oct-1 is a stress sensor. *Cancer Res* **65**, 10750-8 (2005).

97. Ogino, H. & Yasuda, K. [Involvement of maf gene family in crystallin gene regulation]. *Tanpakushitsu Kakusan Koso* **41**, 1050-7 (1996).

98. Kim, J.I., Li, T., Ho, I.C., Grusby, M.J. & Glimcher, L.H. Requirement for the c-Maf transcription factor in crystallin gene regulation and lens development. *Proc Natl Acad Sci U S A* **96**, 3781-5 (1999).

99. Reza, H.M., Ogino, H. & Yasuda, K. L-Maf, a downstream target of Pax6, is essential for chick lens development. *Mech Dev* **116**, 61-73 (2002).

100. Ring, B.Z., Cordes, S.P., Overbeek, P.A. & Barsh, G.S. Regulation of mouse lens fiber cell development and differentiation by the Maf gene. *Development* **127**, 307-17 (2000).

101. Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. & Episkopou, V. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes Dev* **12**, 776-81 (1998).

102. Wyatt, A. *et al.* Novel heterozygous OTX2 mutations and whole gene deletions in anophthalmia, microphthalmia and coloboma. *Hum Mutat* **29**, E278-83 (2008).

103. Blixt, A. *et al.* A forkhead gene, FoxE3, is essential for lens epithelial proliferation and closure of the lens vesicle. *Genes Dev* **14**, 245-54 (2000).

104. Brownell, I., Dirksen, M. & Jamrich, M. Forkhead Foxe3 maps to the dysgenetic lens locus and is critical in lens development and differentiation. *Genesis* **27**, 81-93 (2000).

105. Tholozan, F.M., Sanderson, J.M. & Quinlan, R.A. Focus on molecules: FoxE3. *Exp Eye Res* **84**, 799-800 (2007).

106. Semina, E.V., Brownell, I., Mintz-Hittner, H.A., Murray, J.C. & Jamrich, M.

Mutations in the human forkhead transcription factor FOXE3 associated with anterior segment ocular dysgenesis and cataracts. *Hum Mol Genet* **10**, 231-6 (2001).

107. Butt, T. *et al.* Localization of autosomal recessive congenital cataracts in consanguineous Pakistani families to a new locus on chromosome 1p. *Mol Vis* **13**, 1635-40 (2007).

108. Matsuo, I., Kuratani, S., Kimura, C., Takeda, N. & Aizawa, S. Mouse Otx2 functions in the formation and patterning of rostral head. *Genes Dev* **9**, 2646-58 (1995).

109. Acampora, D. *et al.* Differential transcriptional control as the major molecular event in generating Otx1-/- and Otx2-/- divergent phenotypes. *Development* **126**, 1417-26 (1999).

110. Martinez-Morales, J.R., Signore, M., Acampora, D., Simeone, A. & Bovolenta, P. Otx genes are required for tissue specification in the developing eye. *Development* **128**, 2019-30 (2001).

111. Hettmann, T., Barton, K. & Leiden, J.M. Microphthalmia due to p53-mediated apoptosis of anterior lens epithelial cells in mice lacking the CREB-2 transcription factor. *Dev Biol* **222**, 110-23 (2000).

112. Ohtaka-Maruyama, C., Wang, X., Ge, H. & Chepelinsky, A.B. Overlapping Sp1 and AP2 binding sites in a promoter element of the lens-specific MIP gene. *Nucleic Acids Res* **26**, 407-14 (1998).

113. Semina, E.V., Murray, J.C., Reiter, R., Hrstka, R.F. & Graw, J. Deletion in the promoter region and altered expression of Pitx3 homeobox gene in aphakia mice. *Hum Mol Genet* **9**, 1575-85 (2000).

114. Xu, P.X., Woo, I., Her, H., Beier, D.R. & Maas, R.L. Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. *Development* **124**, 219-31 (1997).

115. Lovicu, F.J., McAvoy, J.W. & de Iongh, R.U. Understanding the role of growth factors in embryonic development: insights from the lens. *Philos Trans R Soc Lond B Biol Sci* **366**, 1204-18 (2011).

116. Wride, M.A., Parker, E. & Sanders, E.J. Members of the bcl-2 and caspase families regulate nuclear degeneration during chick lens fibre differentiation. *Dev Biol* **213**, 142-56 (1999).

117. Fromm, L. & Overbeek, P.A. Inhibition of cell death by lens-specific overexpression of bcl-2 in transgenic mice. *Dev Genet* **20**, 276-87 (1997).

118. Pan, H. & Griep, A.E. Temporally distinct patterns of p53-dependent and p53-independent apoptosis during mouse lens development. *Genes Dev* **9**, 2157-69 (1995).

119. Pokroy, R., Tendler, Y., Pollack, A., Zinder, O. & Weisinger, G. p53 expression in the normal murine eye. *Invest Ophthalmol Vis Sci* **43**, 1736-41 (2002).

120. Weber, G.F. & Menko, A.S. Phosphatidylinositol 3-kinase is necessary for lens fiber cell differentiation and survival. *Invest Ophthalmol Vis Sci* **47**, 4490-9 (2006).

121. Bassnett, S. & McNulty, R. The effect of elevated intraocular oxygen on organelle degradation in the embryonic chicken lens. *J Exp Biol* **206**, 4353-61 (2003).

122. Jun, G. *et al.* EPHA2 is associated with age-related cortical cataract in mice and humans. *PLoS Genet* **5**, e1000584 (2009).

123. Zhang, T. *et al.* Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. *Hum Mutat* **30**, E603-11 (2009).

124. Cooper, M.A. *et al.* Loss of ephrin-A5 function disrupts lens fiber cell packing and leads to cataract. *Proc Natl Acad Sci U S A* **105**, 16620-5 (2008).

125. Truscott, R. Human Age-Related Cataract: A Condition with No Appropriate Animal Model. *Journal of Clinical & Experimental Ophthalmology* **2**, 1 (2011).

126. Truscott, R.J.W. Age-related nuclear cataract—oxidation is the key. *Experimental Eye Research* **80**, 709-725 %U http://linkinghub.elsevier.com/retrieve/pii/S0014483504003549 (2005).

127. Tsentalovich, Y.P. *et al.* Photochemical properties of UV Filter molecules of the human eye. *Invest Ophthalmol Vis Sci* **52**, 7687-96 (2011).

128. Klein, B.E., Klein, R., Lee, K.E., Knudtson, M.D. & Tsai, M.Y. Markers of inflammation, vascular endothelial dysfunction, and age-related cataract. *Am J Ophthalmol* **141**, 116-22 (2006).

129. Sanders, J.L. *et al.* The association of cataract with leukocyte telomere length in older adults: defining a new marker of aging. *J Gerontol A Biol Sci Med Sci* **66**, 639-45 (2011).

130. Donaldson, P.J., Grey, A.C., Maceo Heilman, B., Lim, J.C. & Vaghefi, E. The physiological optics of the lens. *Prog Retin Eye Res* **56**, e1-e24 (2017).

131. Bhatnagar, R. *et al.* Risk of cataract and history of severe diarrheal disease in southern India. *Arch Ophthalmol* **109**, 696-9 (1991).

132. Moffat, B.A., Landman, K.A., Truscott, R.J., Sweeney, M.H. & Pope, J.M. Age-related changes in the kinetics of water transport in normal human lenses. *Exp Eye Res* **69**, 663-9 (1999).

133. Zodpey, S.P., Ughade, S.N., Khanolkar, V.A. & Shrikhande, S.N. Dehydrational crisis from severe diarrhoea and risk of age-related cataract. *J Indian Med Assoc* **97**, 13-5, 24 (1999).

134. Bettelheim, F.A., Lizak, M.J. & Zigler, J.S., Jr. Relaxographic studies of aging normal human lenses. *Exp Eye Res* **75**, 695-702 (2002).

135. Chylack, L.T., Jr., Tung, W. & Harding, R. Sorbitol production in the lens: a means of counteracting glucose-derived osmotic stress. *Ophthalmic Res* **18**, 313-20

(1986).

136. Hashim, Z. & Zarina, S. Osmotic stress induced oxidative damage: possible mechanism of cataract formation in diabetes. *J Diabetes Complications* **26**, 275-9 (2012).

137. Donaldson, P. Reversing the age-dependent decline in lens transport: a new strategy to prevent age related nuclear cataract? *Invest Ophthalmol Vis Sci* **54**, 7188 (2013).

138. Harocopos, G.J. *et al.* Importance of vitreous liquefaction in age-related cataract. *Invest Ophthalmol Vis Sci* **45**, 77-85 (2004).

139. Smuda, M. *et al.* Comprehensive analysis of maillard protein modifications in human lenses: effect of age and cataract. *Biochemistry* **54**, 2500-7 (2015).

140. Bron, A.J., Sparrow, J., Brown, N.A., Harding, J.J. & Blakytny, R. The lens in diabetes. *Eye (Lond)* **7 (Pt 2)**, 260-75 (1993).

141. Qianqian, Y. *et al.* Differential Protein Expression between Type 1 Diabetic Cataract and Age-Related Cataract Patients. *Folia Biol (Praha)* **61**, 74-80 (2015).

142. David, L.L. & Shearer, T.R. Role of proteolysis in lenses: a review. *Lens Eye Toxic Res* **6**, 725-47 (1989).

143. Zetterberg, M. & Celojevic, D. Gender and cataract--the role of estrogen. *Curr Eye Res* **40**, 176-90 (2015).

144. Sharma, K.K. & Santhoshkumar, P. Lens aging: effects of crystallins. *Biochim Biophys Acta* **1790**, 1095-108 (2009).

145. Shang, F. & Taylor, A. Function of the ubiquitin proteolytic pathway in the eye. *Exp Eye Res* **78**, 1-14 (2004).

146. Srivastava, O.P., Srivastava, K. & Silney, C. Levels of crystallin fragments and identification of their origin in water soluble high molecular weight (HMW) proteins of human lenses. *Curr Eye Res* **15**, 511-20 (1996).

147. Gerster, H. Antioxidant vitamins in cataract prevention. *Z Ernahrungswiss* **28**, 56-75 (1989).

148. Zima, T. *et al.* Oxidative stress, metabolism of ethanol and alcohol-related diseases. *J Biomed Sci* **8**, 59-70 (2001).

149. Ogiso, M., Komoto, M., Okinaga, T., Koyota, S. & Hoshi, M. Age-related changes in ganglioside composition in human lens. *Exp Eye Res* **60**, 317-23 (1995).

150. Yappert, M.C., Rujoi, M., Borchman, D., Vorobyov, I. & Estrada, R. Glycero- versus sphingo-phospholipids: correlations with human and non-human mammalian lens growth. *Exp Eye Res* **76**, 725-34 (2003).

151. Gunning, S.J., Chung, K.K., Donaldson, P.J. & Webb, K.F. Identification of a nonselective cation channel in isolated lens fiber cells that is activated by cell

shrinkage. *Am J Physiol Cell Physiol* **303**, C1252-9 (2012).

152. Zeng, J., Borchman, D. & Paterson, C.A. Acute effect of ethanol on lens cation homeostasis. *Alcohol* **16**, 189-93 (1998).

153. Morris, M.S. *et al.* Moderate alcoholic beverage intake and early nuclear and cortical lens opacities. *Ophthalmic Epidemiol* **11**, 53-65 (2004).

154. Tsutsumi, K., Inoue, Y. & Yoshida, C. Acceleration of development of diabetic cataract by hyperlipidemia and low high-density lipoprotein in rats. *Biol Pharm Bull* **22**, 37-41 (1999).

155. Gomez-Ambrosi, J., Salvador, J. & Fruhbeck, G. Is hyperleptinemia involved in the development of age-related lens opacities? *Am J Clin Nutr* **79**, 888-9; author reply 889 (2004).

156. Katsuki, A. *et al.* Increased oxidative stress is associated with serum levels of triglyceride, insulin resistance, and hyperinsulinemia in Japanese metabolically obese, normal-weight men. *Diabetes Care* **27**, 631-2 (2004).

157. Lou, M.F. Redox regulation in the lens. *Prog Retin Eye Res* **22**, 657-82 (2003).

158. Nemet, I. & Monnier, V.M. Vitamin C degradation products and pathways in the human lens. *J Biol Chem* **286**, 37128-36 (2011).

159. Rácz, P. & Erdöhelyi, Á. Cadmium, Lead and Copper Concentrations in Normal and Senile Cataractous Human Lenses. *Ophthalmic Research* **20**, 10-13 %U http://www.karger.com/doi/10.1159/000266248 (1988).

160. Cekic, O. Effect of cigarette smoking on copper, lead, and cadmium accumulation in human lens. *Br J Ophthalmol* **82**, 186-8 (1998).

161. Langford-Smith, A. *et al.* Age and Smoking Related Changes in Metal Ion Levels in Human Lens: Implications for Cataract Formation. *PLoS One* **11**, e0147576 (2016).

162. Zhao, L. *et al.* Vitamin C inhibit the proliferation, migration and epithelial-mesenchymal-transition of lens epithelial cells by destabilizing HIF-1alpha. *Int J Clin Exp Med* **8**, 15155-63 (2015).

163. Weikel, K.A., Garber, C., Baburins, A. & Taylor, A. Nutritional modulation of cataract. *Nutr Rev* **72**, 30-47 (2014).

164. Heiba, I.M., Elston, R.C., Klein, B.E. & Klein, R. Genetic etiology of nuclear cataract: evidence for a major gene. *Am J Med Genet* **47**, 1208-14 (1993).

165. Familial aggregation of lens opacities: the Framingham Eye Study and the Framingham Offspring Eye Study. *Am J Epidemiol* **140**, 555-64 (1994).

166. McCarty, C.A., Mukesh, B.N., Fu, C.L. & Taylor, H.R. The epidemiology of cataract in Australia. *Am J Ophthalmol* **128**, 446-65 (1999).

167. Hammond, C.J., Snieder, H., Spector, T.D. & Gilbert, C.E. Genetic and

Environmental Factors in Age-Related Nuclear Cataracts in Monozygotic and Dizygotic Twins. *New England Journal of Medicine* **342**, 1786-1790 %U http://www.nejm.org/doi/abs/10.1056/NEJM200006153422404 (2000).

168. Congdon, N. *et al.* Nuclear cataract shows significant familial aggregation in an older population after adjustment for possible shared environmental factors. *Invest Ophthalmol Vis Sci* **45**, 2182-6 (2004).

169. Hammond, C.J. *et al.* The heritability of age-related cortical cataract: the twin eye study. *Invest Ophthalmol Vis Sci* **42**, 601-5 (2001).

170. Congdon, N. *et al.* Cortical, but not posterior subcapsular, cataract shows significant familial aggregation in an older population after adjustment for possible shared environmental factors. *Ophthalmology* **112**, 73-7 (2005).

171. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-8 (2012).

172. Pichi, F., Lembo, A., Serafino, M. & Nucci, P. Genetics of Congenital Cataract. *Dev Ophthalmol* **57**, 1-14 (2016).

173. Stambolian, D. *et al.* Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. *Nat Genet* **10**, 307-12 (1995).

174. Kalaydjieva, L. *et al.* A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). *Am J Hum Genet* **65**, 1299-307 (1999).

175. Okano, Y. *et al.* A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant, "Osaka," in Asians. *Am J Hum Genet* **68**, 1036-42 (2001).

176. Kannengiesser, C. *et al.* A new missense mutation in the L ferritin coding sequence associated with elevated levels of glycosylated ferritin in serum and absence of iron overload. *Haematologica* **94**, 335-9 (2009).

177. Bennett, T.M. *et al.* Noncoding variation of the gene for ferritin light chain in hereditary and age-related cataract. *Mol Vis* **19**, 835-44 (2013).

178. Alberti, G. *et al.* Glutathione S-transferase M1 genotype and age-related cataracts. Lack of association in an Italian population. *Invest Ophthalmol Vis Sci* **37**, 1167-73 (1996).

179. Juronen, E. *et al.* Polymorphic glutathione S-transferases as genetic risk factors for senile cortical cataract in Estonians. *Invest Ophthalmol Vis Sci* **41**, 2262-7 (2000).

180. Andersson, M.E. *et al.* Variability in the kinesin light chain 1 gene may influence risk of age-related cataract. *Mol Vis* **13**, 993-6 (2007).

181. Shi, Y. *et al.* Mutation screening of HSF4 in 150 age-related cataract patients. *Mol Vis* **14**, 1850-5 (2008).

182. Pinna, A., Pes, A., Zinellu, A., Carta, A. & Solinas, G. Glucose-6-phosphate dehydrogenase (G6PD) deficiency and senile cataract in a Sardinian male population, Italy. *Ophthalmic Epidemiol* **16**, 395-9 (2009).

183. Zhou, J., Hu, J. & Guan, H. The association between copy number variations in glutathione S-transferase M1 and T1 and age-related cataract in a Han Chinese population. *Invest Ophthalmol Vis Sci* **51**, 3924-8 (2010).

184. Zuercher, J. *et al.* Alterations of the 5'untranslated region of SLC16A12 lead to age-related cataract. *Invest Ophthalmol Vis Sci* **51**, 3354-61 (2010).

185. Mamata, M., Sridhar, G., Reddy, K.R., Nagaraju, T. & Padma, T. Is the variant c. 422+90G --> A in intron 4 of indoleamine 2, 3 -dioxygenase (IDO) gene related to age related cataracts? *Mol Vis* **17**, 1203-8 (2011).

186. Padma, G., Mamata, M., Reddy, K.R. & Padma, T. Polymorphisms in two DNA repair genes (XPD and XRCC1)--association with age related cataracts. *Mol Vis* **17**, 127-33 (2011).

187. Zhang, Y. *et al.* Genetic polymorphisms of superoxide dismutases, catalase, and glutathione peroxidase in age-related cataract. *Mol Vis* **17**, 2325-32 (2011).

188. Zhou, Z. *et al.* Major intrinsic protein (MIP) polymorphism is associated with age-related cataract in Chinese. *Mol Vis* **17**, 2292-6 (2011).

189. Sadikov, T. *et al.* C1824T mutation in the LMNA gene has no association with senile cataract. *Neurobiol Aging* **33**, 1487 e15-9 (2012).

190. Zhang, Y. *et al.* Genetic polymorphisms in DNA repair genes OGG1, APE1, XRCC1, and XPD and the risk of age-related cataract. *Ophthalmology* **119**, 900-6 (2012).

191. Ji, F., Jung, J., Koharudin, L.M. & Gronenborn, A.M. The human W42R gammaD-crystallin mutant structure provides a link between congenital and age-related cataracts. *J Biol Chem* **288**, 99-109 (2013).

192. Jiang, J. *et al.* Copy number variations of DNA repair genes and the age-related cataract: Jiangsu Eye Study. *Invest Ophthalmol Vis Sci* **54**, 932-8 (2013).

193. Jiang, S. *et al.* Polymorphisms of the WRN gene and DNA damage of peripheral lymphocytes in age-related cataract in a Han Chinese population. *Age (Dordr)* **35**, 2435-44 (2013).

194. Lin, Q. *et al.* Genetic variations and polymorphisms in the ezrin gene are associated with age-related cataract. *Mol Vis* **19**, 1572-9 (2013).

195. Zhang, L. *et al.* Association of a rare haplotype in Kinesin light chain 1 gene with age-related cataract in a han chinese population. *PLoS One* **8**, e64052 (2013).

196. Zhang, Y. *et al.* Genetic polymorphisms of HSP70 in age-related cataract. *Cell Stress Chaperones* **18**, 703-9 (2013).

197. Lin, Q., Zhou, N., Zhang, N. & Qi, Y. Mutational screening of EFNA5 in Chinese age-related cataract patients. *Ophthalmic Res* **52**, 124-9 (2014).

198. Zheng, C. *et al.* RNA granule component TDRD7 gene polymorphisms in a Han Chinese population with age-related cataract. *J Int Med Res* **42**, 153-63 (2014).

199. Skiljic, D. *et al.* Estrogen-related Polymorphisms in Estonian Patients with Age-related Cataract. *Ophthalmic Genet* **36**, 188-91 (2015).

200. Sun, W., Su, L., Sheng, Y., Shen, Y. & Chen, G. Is there association between Glutathione S Transferases polymorphisms and cataract risk: a meta-analysis? *BMC Ophthalmol* **15**, 84 (2015).

201. Wang, C., Lai, Q., Zhang, S. & Hu, J. Senile cataract and genetic polymorphisms of APE1, XRCC1 and OGG1. *Int J Clin Exp Pathol* **8**, 16036-45 (2015).

202. Wang, X.B. *et al.* Associations of Polymorphisms in MTHFR Gene with the Risk of Age-Related Cataract in Chinese Han Population: A Genotype-Phenotype Analysis. *PLoS One* **10**, e0145581 (2015).

203. Wu, M. *et al.* The link between apolipoprotein E, presenilin 1, and kinesin light chain 1 gene polymorphisms and age-related cortical cataracts in the Chinese population. *Mol Vis* **21**, 412-6 (2015).

204. Ma, X., Jiao, X., Ma, Z. & Hejtmancik, J.F. Polymorphism rs7278468 is associated with Age-related cataract through decreasing transcriptional activity of the CRYAA promoter. *Sci Rep* **6**, 23206 (2016).

205. Cui, N.H., Qiao, C., Chang, X.K. & Wei, L. Associations of PARP-1 variant rs1136410 with PARP activities, oxidative DNA damage, and the risk of age-related cataract in a Chinese Han population: A two-stage case-control analysis. *Gene* **600**, 70-76 (2017).

206. Eiberg, H., Lund, A.M., Warburg, M. & Rosenberg, T. Assignment of congenital cataract Volkmann type (CCV) to chromosome 1p36. *Hum Genet* **96**, 33-8 (1995).

207. McKay, J.D. *et al.* The telomere of human chromosome 1p contains at least two independent autosomal dominant congenital cataract genes. *Br J Ophthalmol* **89**, 831-4 (2005).

208. Tan, W. *et al.* Association of EPHA2 polymorphisms and age-related cortical cataract in a Han Chinese population. *Mol Vis* **17**, 1553-8 (2011).

209. Sundaresan, P. *et al.* EPHA2 polymorphisms and age-related cataract in India. *PLoS One* **7**, e33001 (2012).

210. Yang, J. *et al.* Association of the ephreceptor tyrosinekinase-type A2 (EPHA2) gene polymorphism rs3754334 with age-related cataract risk: a meta-analysis. *PLoS One* **8**, e71003 (2013).

211. Zhang, G. *et al.* EphA2 is an essential mediator of UV radiation-induced

apoptosis. *Cancer Res* **68**, 1691-6 (2008).

212. Udayakumar, D. *et al.* EphA2 is a critical oncogene in melanoma. *Oncogene* **30**, 4921-9 (2011).

213. Mackay, D.S., Andley, U.P. & Shiels, A. Cell death triggered by a novel mutation in the alphaA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. *Eur J Hum Genet* **11**, 784-93 (2003).

214. Wistow, G. The human crystallin gene families. *Hum Genomics* **6**, 26 (2012).

215. England, S.K., Uebele, V.N., Kodali, J., Bennett, P.B. & Tamkun, M.M. A novel K+ channel beta-subunit (hKv beta 1.3) is produced via alternative mRNA splicing. *J Biol Chem* **270**, 28531-4 (1995).

216. Eom, S.Y. *et al.* A pilot exome-wide association study of age-related cataract in Koreans. *J Biomed Res* **30**, 186-90 (2016).

217. Ritchie, M.D. *et al.* Electronic medical records and genomics (eMERGE) network exploration in cataract: several new potential susceptibility loci. *Mol Vis* **20**, 1281-95 (2014).

218. Hall, M.A. *et al.* Biology-Driven Gene-Gene Interaction Analysis of Age-Related Cataract in the eMERGE Network. *Genet Epidemiol* **39**, 376-84 (2015).

219. Pendergrass, S.A. *et al.* Next-generation analysis of cataracts: determining knowledge driven gene-gene interactions using biofilter, and gene-environment interactions using the Phenx Toolkit*. *Pac Symp Biocomput*, 495-505 (2015).

220. Mundy, K.M., Nichols, E. & Lindsey, J. Socioeconomic Disparities in Cataract Prevalence, Characteristics, and Management. *Semin Ophthalmol* **31**, 358-63 (2016).

221. Klein, B.E., Klein, R. & Linton, K.L. Prevalence of age-related lens opacities in a population. The Beaver Dam Eye Study. *Ophthalmology* **99**, 546-552 (1992).

222. Leske, M.C., Connell, A.M., Wu, S.Y., Hyman, L. & Schachat, A. Prevalence of lens opacities in the Barbados Eye Study. *Arch Ophthalmol* **115**, 105-11 (1997).

223. Mitchell, P., Cumming, R.G., Attebo, K. & Panchapakesan, J. Prevalence of cataract in Australia: the Blue Mountains eye study. *Ophthalmology* **104**, 581-588 (1997).

224. Congdon, N. *et al.* Prevalence of the different types of age-related cataract in an African population. *Invest Ophthalmol Vis Sci* **42**, 2478-82 (2001).

225. Tsai, S.Y., Hsu, W.M., Cheng, C.Y., Liu, J.H. & Chou, P. Epidemiologic study of age-related cataracts among an elderly Chinese population in Shih-Pai, Taiwan. *Ophthalmology* **110**, 1089-95 (2003).

226. Xu, L. *et al.* Prevalence and risk factors of lens opacities in urban and rural Chinese in Beijing. *Ophthalmology* **113**, 747-55 (2006).

227. Vashist, P. *et al.* Prevalence of cataract in an older population in India: the India

study of age-related eye disease. *Ophthalmology* **118**, 272-8 e1-2 (2011).

228. Storey, P., Munoz, B., Friedman, D. & West, S. Racial differences in lens opacity incidence and progression: the Salisbury Eye Evaluation (SEE) study. *Invest Ophthalmol Vis Sci* **54**, 3010-8 (2013).

229. Mahdi, A.M. *et al.* Prevalence and risk factors for lens opacities in Nigeria: results of the national blindness and low vision survey. *Invest Ophthalmol Vis Sci* **55**, 2642-51 (2014).

230. Varma, R. *et al.* Prevalence of Lens Opacities in Adult Chinese Americans: The Chinese American Eye Study (CHES). *Invest Ophthalmol Vis Sci* **57**, 6692-6699 (2016).

231. Klein, B.E., Klein, R. & Lee, K.E. Incidence of age-related cataract: the Beaver Dam Eye Study. *Arch Ophthalmol* **116**, 219-25 (1998).

232. Leske, M.C. *et al.* Incidence and progression of lens opacities in the Barbados Eye Studies. *Ophthalmology* **107**, 1267-73 (2000).

233. McCarty, C.A., Mukesh, B.N., Dimitrov, P.N. & Taylor, H.R. Incidence and progression of cataract in the Melbourne Visual Impairment Project. *Am J Ophthalmol* **136**, 10-7 (2003).

234. Chandrasekaran, S., Cumming, R.G., Rochtchina, E. & Mitchell, P. Associations between elevated intraocular pressure and glaucoma, use of glaucoma medications, and 5-year incident cataract: the Blue Mountains Eye Study. *Ophthalmology* **113**, 417-24 (2006).

235. Kanthan, G.L. *et al.* Ten-year incidence of age-related cataract and cataract surgery in an older Australian population. The Blue Mountains Eye Study. *Ophthalmology* **115**, 808-814 e1 (2008).

236. Varma, R. *et al.* Four-year incidence and progression of age-related macular degeneration: the Los Angeles Latino Eye Study. *Am J Ophthalmol* **149**, 741-51 (2010).

237. Varma, R. *et al.* Four-year incidence and progression of lens opacities: the Los Angeles Latino Eye Study. *Am J Ophthalmol* **149**, 728-34 e1-2 (2010).

238. Kuang, T.M. *et al.* Seven-year incidence of age-related cataracts among an elderly Chinese population in Shihpai, Taiwan: The Shihpai Eye Study. *Invest Ophthalmol Vis Sci* **54**, 6409-15 (2013).

239. Jonas, J.B., Xu, L. & Wang, Y.X. The Beijing Eye Study. *Acta Ophthalmologica* **87**, 247-261 %U http://doi.wiley.com/10.1111/j.1755-3768.2008.01385.x (2009).

240. Moreau, K.L. & King, J.A. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol Med* **18**, 273-82 (2012).

241. Chua, J. *et al.* Ancestry, Socioeconomic Status, and Age-Related Cataract in Asians: The Singapore Epidemiology of Eye Diseases Study. *Ophthalmology* **122**,

2169-78 (2015).

242. Finger, R.P. *et al.* Migration study of lens opacities in Bangladeshi adults in London and Bangladesh: a pilot study. *Br J Ophthalmol* **99**, 762-7 (2015).

243. Zhang, J.S. *et al.* Five-year incidence of age-related cataract and cataract surgery in the adult population of greater Beijing: the Beijing Eye Study. *Ophthalmology* **118**, 711-8 (2011).

244. Leske, M.C., Wu, S.Y., Connell, A.M., Hyman, L. & Schachat, A.P. Lens opacities, demographic factors and nutritional supplements in the Barbados Eye Study. *Int J Epidemiol* **26**, 1314-22 (1997).

245. Klein, B.E., Klein, R. & Ritter, L.L. Is there evidence of an estrogen effect on age-related lens opacities? The Beaver Dam Eye Study. *Arch Ophthalmol* **112**, 85-91 (1994).

246. Freeman, E.E., Munoz, B., Schein, O.D. & West, S.K. Hormone replacement therapy and lens opacities: the Salisbury Eye Evaluation project. *Arch Ophthalmol* **119**, 1687-92 (2001).

247. Weintraub, J.M. *et al.* Postmenopausal hormone use and lens opacities. *Ophthalmic Epidemiol* **9**, 179-90 (2002).

248. Younan, C. *et al.* Hormone replacement therapy, reproductive factors, and the incidence of cataract and cataract surgery: the Blue Mountains Eye Study. *Am J Epidemiol* **155**, 997-1006 (2002).

249. Kanthan, G.L. *et al.* Exogenous oestrogen exposure, female reproductive factors and the long-term incidence of cataract: the Blue Mountains Eye Study. *Acta Ophthalmol* **88**, 773-8 (2010).

250. Lai, K. *et al.* The effects of postmenopausal hormone use on cataract: a meta-analysis. *PLoS One* **8**, e78647 (2013).

251. Na, K.S. *et al.* The ocular benefits of estrogen replacement therapy: a population-based study in postmenopausal Korean women. *PLoS One* **9**, e106473 (2014).

252. Tian, Y. *et al.* Parity and the risk of cataract: a cross-sectional analysis in the Dongfeng-Tongji cohort study. *Br J Ophthalmol* **99**, 1650-4 (2015).

253. Mascitelli, L., Pezzetta, F. & Sullivan, J.L. Why women live longer than men: sex differences in longevity. *Gend Med* **3**, 341; author reply 342 (2006).

254. Vitale, C., Fini, M., Speziale, G. & Chierchia, S. Gender differences in the cardiovascular effects of sex hormones. *Fundam Clin Pharmacol* **24**, 675-85 (2010).

255. Moisan, F. *et al.* Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis. *J Neurol Neurosurg Psychiatry* **87**, 952-7 (2016).

256. Novella, S., Dantas, A.P., Segarra, G., Medina, P. & Hermenegildo, C. Vascular Aging in Women: is Estrogen the Fountain of Youth? *Front Physiol* **3**, 165 (2012).

257. Keating, N.L., O'Malley, A.J. & Smith, M.R. Diabetes and cardiovascular disease during androgen deprivation therapy for prostate cancer. *J Clin Oncol* **24**, 4448-56 (2006).

258. Beebe-Dimmer, J. *et al.* Androgen deprivation therapy and cataract incidence among elderly prostate cancer patients in the United States. *Ann Epidemiol* **21**, 156-63 (2011).

259. Wickham, L.A. *et al.* Identification of androgen, estrogen and progesterone receptor mRNAs in the eye. *Acta Ophthalmol Scand* **78**, 146-53 (2000).

260. Kirker, M.R., Gallagher, K.M., Witt-Enderby, P.A. & Davis, V.L. High affinity nuclear and nongenomic estradiol binding sites in the human and mouse lens. *Exp Eye Res* **112**, 1-9 (2013).

261. Moosmann, B. & Behl, C. The antioxidant neuroprotective effects of estrogens and phenolic compounds are independent from their estrogenic properties. *Proc Natl Acad Sci U S A* **96**, 8867-72 (1999).

262. Alonso-Alvarez, C., Bertrand, S., Faivre, B., Chastel, O. & Sorci, G. Testosterone and oxidative stress: the oxidation handicap hypothesis. *Proc Biol Sci* **274**, 819-25 (2007).

263. Gajjar, D. *et al.* Rapid action of oestradiol against hydrogen peroxide-induced oxidative stress in cataractous lens epithelium: an in vitro study. *Eye (Lond)* **23**, 1456-63 (2009).

264. Celojevic, D., Petersen, A., Karlsson, J.O., Behndig, A. & Zetterberg, M. Effects of 17beta-estradiol on proliferation, cell viability and intracellular redox status in native human lens epithelial cells. *Mol Vis* **17**, 1987-96 (2011).

265. Moor, A.N., Flynn, J.M., Gottipati, S., Giblin, F.J. & Cammarata, P.R. 17beta-estradiol stimulates MAPK signaling pathway in human lens epithelial cell cultures preventing collapse of mitochondrial membrane potential during acute oxidative stress. *Mitochondrion* **5**, 235-47 (2005).

266. Flynn, J.M., Lannigan, D.A., Clark, D.E., Garner, M.H. & Cammarata, P.R. RNA suppression of ERK2 leads to collapse of mitochondrial membrane potential with acute oxidative stress in human lens epithelial cells. *Am J Physiol Endocrinol Metab* **294**, E589-99 (2008).

267. Hales, A.M., Chamberlain, C.G., Murphy, C.R. & McAvoy, J.W. Estrogen protects lenses against cataract induced by transforming growth factor-beta (TGFbeta). *J Exp Med* **185**, 273-80 (1997).

268. Bigsby, R.M. *et al.* Ovarian hormone modulation of radiation-induced

cataractogenesis: dose-response studies. *Invest Ophthalmol Vis Sci* **50**, 3304-10 (2009).

269. Bigsby, R.M., Cardenas, H., Caperell-Grant, A. & Grubbs, C.J. Protective effects of estrogen in a rat model of age-related cataracts. *Proc Natl Acad Sci U S A* **96**, 9328-32 (1999).

270. Harding, J.J. & van Heyningen, R. Drugs, including alcohol, that act as risk factors for cataract, and possible protection against cataract by aspirin-like analgesics and cyclopenthiazide. *Br J Ophthalmol* **72**, 809-14 (1988).

271. Hankinson, S.E. *et al.* A prospective study of cigarette smoking and risk of cataract surgery in women. *JAMA* **268**, 994-8 (1992).

272. Leske, M.C., Chylack, L.T., Jr. & Wu, S.Y. The Lens Opacities Case-Control Study. Risk factors for cataract. *Arch Ophthalmol* **109**, 244-51 (1991).

273. Age-Related Eye Disease Study Research, G. Risk factors associated with age-related nuclear and cortical cataract : a case-control study in the Age-Related Eye Disease Study, AREDS Report No. 5. *Ophthalmology* **108**, 1400-8 (2001).

274. Theodoropoulou, S. *et al.* The epidemiology of cataract: a study in Greece. *Acta Ophthalmol* **89**, e167-73 (2011).

275. Christen, W.G. *et al.* A prospective study of cigarette smoking and risk of cataract in men. *JAMA* **268**, 989-93 (1992).

276. Arnarsson, A. *et al.* Risk factors for nuclear lens opacification: the Reykjavik Eye Study. *Dev Ophthalmol* **35**, 12-20 (2002).

277. Ye, J. *et al.* Smoking and risk of age-related cataract: a meta-analysis. *Invest Ophthalmol Vis Sci* **53**, 3885-95 (2012).

278. Flaye, D.E., Sullivan, K.N., Cullinan, T.R., Silver, J.H. & Whitelocke, R.A. Cataracts and cigarette smoking. The City Eye Study. *Eye (Lond)* **3 (Pt 4)**, 379-84 (1989).

279. Hiller, R. *et al.* Cigarette smoking and the risk of development of lens opacities. The Framingham studies. *Arch Ophthalmol* **115**, 1113-8 (1997).

280. Christen, W.G. *et al.* Smoking cessation and risk of age-related cataract in men. *JAMA* **284**, 713-6 (2000).

281. West, S. *et al.* Cigarette smoking and risk for progression of nuclear opacities. *Arch Ophthalmol* **113**, 1377-80 (1995).

282. Ojofeitimi, E.O. *et al.* Dietary and lifestyle patterns in the aetiology of cataracts in Nigerian patients. *Nutr Health* **13**, 61-8 (1999).

283. Richter, G.M. *et al.* Risk factors for cortical, nuclear, posterior subcapsular, and mixed lens opacities: the Los Angeles Latino Eye Study. *Ophthalmology* **119**, 547-54 (2012).

284. Foster, P.J., Wong, T.Y., Machin, D., Johnson, G.J. & Seah, S.K. Risk factors for nuclear, cortical and posterior subcapsular cataracts in the Chinese population of Singapore: the Tanjong Pagar Survey. *Br J Ophthalmol* **87**, 1112-20 (2003).

285. Krishnaiah, S. *et al.* Smoking and its association with cataract: results of the Andhra Pradesh eye disease study from India. *Invest Ophthalmol Vis Sci* **46**, 58-65 (2005).

286. Mamatha, B.S., Nidhi, B., Padmaprabhu, C.A., Pallavi, P. & Vallikannan, B. Risk Factors for Nuclear and Cortical Cataracts: A Hospital Based Study. *J Ophthalmic Vis Res* **10**, 243-9 (2015).

287. Wu, R. *et al.* Smoking, socioeconomic factors, and age-related cataract: The Singapore Malay Eye study. *Arch Ophthalmol* **128**, 1029-35 (2010).

288. Halliwell, B. & Poulsen, H.E. *Cigarette smoke and oxidative stress*, xi, 407 p. (Springer, Berlin ; New York, 2006).

289. Ramakrishnan, S. *et al.* Smoking of beedies and cataract: cadmium and vitamin C in the lens and blood. *Br J Ophthalmol* **79**, 202-6 (1995).

290. Avunduk, A.M., Yardimci, S., Avunduk, M.C., Kurnaz, L. & Kockar, M.C. Determinations of some trace and heavy metals in rat lenses after tobacco smoke exposure and their relationships to lens injury. *Exp Eye Res* **65**, 417-23 (1997).

291. Kalariya, N.M., Nair, B., Kalariya, D.K., Wills, N.K. & van Kuijk, F.J. Cadmium-induced induction of cell death in human lens epithelial cells: implications to smoking associated cataractogenesis. *Toxicol Lett* **198**, 56-62 (2010).

292. Mosad, S.M. *et al.* Lens cadmium, lead, and serum vitamins C, E, and beta carotene in cataractous smoking patients. *Curr Eye Res* **35**, 23-30 (2010).

293. Li, B. *et al.* Relationship Between the Altered Expression and Epigenetics of GSTM3 and Age-Related Cataract. *Invest Ophthalmol Vis Sci* **57**, 4721-32 (2016).

294. Wang, Y. & Guan, H. The Role of DNA Methylation in Lens Development and Cataract Formation. *Cell Mol Neurobiol* **37**, 979-984 (2017).

295. Wang, Y., Zhang, G., Kang, L. & Guan, H. Expression Profiling of DNA Methylation and Transcriptional Repression Associated Genes in Lens Epithelium Cells of Age-Related Cataract. *Cell Mol Neurobiol* **37**, 537-543 (2017).

296. General, T.S. Chemistry and Toxicology of Cigarette Smoke and Biomarkers of Exposure and Harm. in How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General (Atlanta (GA), 2010).

297. Gao, X., Jia, M., Zhang, Y., Breitling, L.P. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics* **7**, 113 (2015).

298. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**, 436-447 (2016).

299. Marabita, F. *et al.* Smoking induces DNA methylation changes in Multiple Sclerosis patients with exposure-response relationship. *Sci Rep* **7**, 14589 (2017).

300. Wojno, T., Singer, D. & Schultz, R.O. Ultraviolet light, cataracts, and spectacle wear. *Ann Ophthalmol* **15**, 729-32 (1983).

301. Bochow, T.W. *et al.* Ultraviolet light exposure and risk of posterior subcapsular cataracts. *Arch Ophthalmol* **107**, 369-72 (1989).

302. Wong, L. *et al.* Sunlight exposure, antioxidant status, and cataract in Hong Kong fishermen. *J Epidemiol Community Health* **47**, 46-9 (1993).

303. Burton, M. *et al.* The prevalence of cataract in two villages of northern Pakistan with different levels of ultraviolet radiation. *Eye (London, England)* **11 (Pt 1)**, 95 (1997).

304. Neale, R.E., Purdie, J.L., Hirst, L.W. & Green, A.C. Sun exposure as a risk factor for nuclear cataract. *Epidemiology* **14**, 707-12 (2003).

305. Hiller, R., Giacometti, L. & Yuen, K. Sunlight and cataract: an epidemiologic investigation. *Am J Epidemiol* **105**, 450-9 (1977).

306. Javitt, J.C. & Taylor, H.R. Cataract and latitude. *Doc Ophthalmol* **88**, 307-25 (1994).

307. Delcourt, C. *et al.* Lifetime exposure to ambient ultraviolet radiation and the risk for cataract extraction and age-related macular degeneration: the Alienor Study. *Invest Ophthalmol Vis Sci* **55**, 7619-27 (2014).

308. Perkins, E.S. The association between pinguecula, sunlight and cataract. *Ophthalmic Res* **17**, 325-30 (1985).

309. Collman, G.W., Shore, D.L., Shy, C.M., Checkoway, H. & Luria, A.S. Sunlight and other risk factors for cataracts: an epidemiologic study. *Am J Public Health* **78**, 1459-62 (1988).

310. Dolezal, J.M., Perkins, E.S. & Wallace, R.B. Sunlight, skin sensitivity, and senile cataract. *Am J Epidemiol* **129**, 559-68 (1989).

311. Pastor-Valero, M., Fletcher, A.E., de Stavola, B.L. & Chaques-Alepuz, V. Years of sunlight exposure and cataract: a case-control study in a Mediterranean population. *BMC Ophthalmol* **7**, 18 (2007).

312. Katoh, N. *et al.* Cortical lens opacification in Iceland. Risk factor analysis -- Reykjavik Eye Study. *Acta Ophthalmol Scand* **79**, 154-9 (2001).

313. Hirvela, H., Luukinen, H. & Laatikainen, L. Prevalence and risk factors of lens opacities in the elderly in Finland. A population-based study. *Ophthalmology* **102**, 108-17 (1995).

314. Taylor, H.R. The environment and the lens. *Br J Ophthalmol* **64**, 303-10 (1980).

315. Hollows, F. & Moran, D. Cataract--the ultraviolet risk factor. *Lancet* **2**, 1249-50 (1981).

316. McCarty, C.A., Nanjan, M.B. & Taylor, H.R. Attributable risk estimates for cataract to prioritize medical and public health action. *Invest Ophthalmol Vis Sci* **41**, 3720-5 (2000).

317. Taylor, H.R. *et al.* Effect of ultraviolet radiation on cataract formation. *N Engl J Med* **319**, 1429-33 (1988).

318. Cruickshanks, K.J., Klein, B.E. & Klein, R. Ultraviolet light exposure and lens opacities: the Beaver Dam Eye Study. *Am J Public Health* **82**, 1658-62 (1992).

319. Rosmini, F. *et al.* A dose-response effect between a sunlight index and age-related cataracts. Italian-American Cataract Study Group. *Ann Epidemiol* **4**, 266-70 (1994).

320. West, S., Duncan, D., Munoz, B. & Rubin, G. Sunlight exposure and risk of lens opacities in a population-based study: The Salisbury Eye Evaluation Project. *JAMA* **280**, 714-8 (1998).

321. Delcourt, C. *et al.* Light exposure and the risk of cortical, nuclear, and posterior subcapsular cataracts: the Pathologies Oculaires Liees a l'Age (POLA) study. *Arch Ophthalmol* **118**, 385-92 (2000).

322. Tang, Y. *et al.* The Association of Outdoor Activity and Age-Related Cataract in a Rural Population of Taizhou Eye Study: Phase 1 Report. *PLoS One* **10**, e0135870 (2015).

323. Vines, A.P. An epidemiological sample survey of the Highlands, Mainland and Islands Regions of the Territory of Papua and New Guinea, (Port Moresby : Bloink, Port Moresby, 1970).

324. Goldsmith, R.I., Rothhammer, F. & Schull, W.J. The multinational Andean genetic and health program: III. Ophthalmic disease and disability among the Aymara. *Bull Pan Am Health Organ* **13**, 58-65 (1979).

325. Song, W.X. [A survey of senile cataract among Tibetans in Chang-Du District, Tibet (author's transl)]. *Zhonghua Yan Ke Za Zhi* **15**, 100-4 (1979).

326. Shrestha, S., Shrestha, S.M. & Gurung, A. Comparative Study of Prevalence of Cataract at High Altitude and Kathmandu Valley. *J Nepal Health Res Counc* **14**, 81-84 (2016).

327. Yu, J.M. *et al.* Prevalence and risk factors of lens opacities in rural populations living at two different altitudes in China. *Int J Ophthalmol* **9**, 610-6 (2016).

328. Chatterjee, A., Milton, R.C. & Thyle, S. Prevalence and aetiology of cataract in Punjab. *Br J Ophthalmol* **66**, 35-42 (1982).

329. Brilliant, L.B. *et al.* Associations among cataract prevalence, sunlight hours, and

altitude in the Himalayas. *Am J Epidemiol* **118**, 250-64 (1983).

330. Mohan, M. *et al.* India-US case-control study of age-related cataracts. India-US Case-Control Study Group. *Arch Ophthalmol* **107**, 670-6 (1989).

331. Beall, C.M. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integr Comp Biol* **46**, 18-24 (2006).

332. Bigham, A.W. Genetics of human origin and evolution: high-altitude adaptations. *Curr Opin Genet Dev* **41**, 8-13 (2016).

333. Lofgren, S. Solar ultraviolet radiation cataract. *Exp Eye Res* **156**, 112-116 (2017).

334. Zhang, J., Yan, H., Lofgren, S., Tian, X. & Lou, M.F. Ultraviolet radiation-induced cataract in mice: the effect of age and the potential biochemical mechanism. *Invest Ophthalmol Vis Sci* **53**, 7276-85 (2012).

335. Moran, S.D., Zhang, T.O., Decatur, S.M. & Zanni, M.T. Amyloid fiber formation in human gammaD-Crystallin induced by UV-B photodamage. *Biochemistry* **52**, 6169-81 (2013).

336. Linetsky, M. *et al.* UVA light-excited kynurenines oxidize ascorbate and modify lens proteins through the formation of advanced glycation end products: implications for human lens aging and cataract formation. *J Biol Chem* **289**, 17111-23 (2014).

337. Chauss, D., Brennan, L.A., Bakina, O. & Kantorow, M. Integrin alphaVbeta5-mediated Removal of Apoptotic Cell Debris by the Eye Lens and Its Inhibition by UV Light Exposure. *J Biol Chem* **290**, 30253-66 (2015).

338. Dong, Y., Zheng, Y., Xiao, J., Zhu, C. & Zhao, M. Regulatory effect of Bcl-2 in ultraviolet radiation-induced apoptosis of the mouse crystalline lens. *Exp Ther Med* **11**, 973-977 (2016).

339. Wu, Q., Guo, D., Bi, H., Wang, D. & Du, Y. UVB irradiation-induced dysregulation of plasma membrane calcium ATPase1 and intracellular calcium homeostasis in human lens epithelial cells. *Mol Cell Biochem* **382**, 263-72 (2013).

340. Meyer, L.M., Lofgren, S., Holz, F.G., Wegener, A. & Soderberg, P. Bilateral cataract induced by unilateral UVR-B exposure -- evidence for an inflammatory response. *Acta Ophthalmol* **91**, 236-42 (2013).

341. Tavani, A., Negri, E. & La Vecchia, C. Food and nutrient intake and risk of cataract. *Annals of Epidemiology* **6**, 41-46 (1996).

342. Cumming, R.G., Mitchell, P. & Smith, W. Dietary sodium intake and cataract: the Blue Mountains Eye Study. *Am J Epidemiol* **151**, 624-6 (2000).

343. Cumming, R.G., Mitchell, P. & Smith, W. Diet and cataract: the Blue Mountains Eye Study. *Ophthalmology* **107**, 450-6 (2000).

344. Ghanavati, M. *et al.* Healthy Eating Index in Patients With Cataract: A Case-

Control Study. *Iran Red Crescent Med J* **17**, e22490 (2015).

345. Jee, D. & Kim, E.C. Association between serum 25-hydroxyvitamin D levels and age-related cataracts. *J Cataract Refract Surg* **41**, 1705-15 (2015).

346. Klein, B.E. *et al.* Supplements and age-related eye conditions the beaver dam eye study. *Ophthalmology* **115**, 1203-8 (2008).

347. Cumming, R.G. & Mitchell, P. Alcohol, smoking, and cataracts: the Blue Mountains Eye Study. *Arch Ophthalmol* **115**, 1296-303 (1997).

348. Klein, B.E., Klein, R., Lee, K.E. & Meuer, S.M. Socioeconomic and lifestyle factors and the 10-year incidence of age-related cataracts. *Am J Ophthalmol* **136**, 506-12 (2003).

349. Li, Z. *et al.* Alcohol consumption and visual impairment in a rural Northern Chinese population. *Ophthalmic Epidemiol* **21**, 384-90 (2014).

350. Lindblad, B.E., Hakansson, N., Philipson, B. & Wolk, A. Alcohol consumption and risk of cataract extraction: a prospective cohort study of women. *Ophthalmology* **114**, 680-5 (2007).

351. Manson, J.E., Christen, W.G., Seddon, J.M., Glynn, R.J. & Hennekens, C.H. A prospective study of alcohol consumption and risk of cataract. *Am J Prev Med* **10**, 156-61 (1994).

352. Chasan-Taber, L. *et al.* A prospective study of alcohol consumption and cataract extraction among U.S. women. *Ann Epidemiol* **10**, 347-53 (2000).

353. Phillips, C.I. *et al.* Human cataract risk factors: significance of abstention from, and high consumption of, ethanol (U-curve) and non-significance of smoking. *Ophthalmic Res* **28**, 237-47 (1996).

354. Wang, W. & Zhang, X. Alcohol intake and the risk of age-related cataracts: a meta-analysis of prospective cohort studies. *PLoS One* **9**, e107820 (2014).

355. Gong, Y., Feng, K., Yan, N., Xu, Y. & Pan, C.W. Different amounts of alcohol consumption and cataract: a meta-analysis. *Optom Vis Sci* **92**, 471-9 (2015).

356. Ai, Y. *et al.* A mouse model of galactose-induced cataracts. *Hum Mol Genet* **9**, 1821-7 (2000).

357. Chiu, C.J. *et al.* Carbohydrate intake and glycemic index in relation to the odds of early cortical and nuclear lens opacities. *Am J Clin Nutr* **81**, 1411-6 (2005).

358. Chiu, C.J., Milton, R.C., Gensler, G. & Taylor, A. Dietary carbohydrate intake and glycemic index in relation to cortical and nuclear lens opacities in the Age-Related Eye Disease Study. *Am J Clin Nutr* **83**, 1177-84 (2006).

359. Lu, Z.Q. *et al.* Relationship between dietary macronutrient intake and the risk of age-related cataract in middle-aged and elderly patients in northeast China. *Int J Ophthalmol* **6**, 805-10 (2013).

360. Tan, J. *et al.* Carbohydrate nutrition, glycemic index, and the 10-y incidence of cataract. *Am J Clin Nutr* **86**, 1502-8 (2007).

361. Townend, B.S. *et al.* Dietary macronutrient intake and five-year incident cataract: the blue mountains eye study. *Am J Ophthalmol* **143**, 932-939 (2007).

362. Theodoropoulou, S. *et al.* Diet and cataract: a case-control study. *Int Ophthalmol* **34**, 59-68 (2014).

363. Wu, H. *et al.* Association between dietary carbohydrate intake and dietary glycemic index and risk of age-related cataract: a meta-analysis. *Invest Ophthalmol Vis Sci* **55**, 3660-8 (2014).

364. Appleby, P.N., Allen, N.E. & Key, T.J. Diet, vegetarianism, and cataract risk. *Am J Clin Nutr* **93**, 1128-35 (2011).

365. Ludwig, D.S. The glycemic index: physiological mechanisms relating to obesity, diabetes, and cardiovascular disease. *JAMA* **287**, 2414-23 (2002).

366. Wolever, T.M., Jenkins, D.J., Jenkins, A.L. & Josse, R.G. The glycemic index: methodology and clinical implications. *Am J Clin Nutr* **54**, 846-54 (1991).

367. Chiu, C.J. *et al.* Dietary carbohydrate in relation to cortical and nuclear lens opacities in the melbourne visual impairment project. *Invest Ophthalmol Vis Sci* **51**, 2897-905 (2010).

368. Schaumberg, D.A., Liu, S., Seddon, J.M., Willett, W.C. & Hankinson, S.E. Dietary glycemic load and risk of age-related cataract. *Am J Clin Nutr* **80**, 489-95 (2004).

369. Turati, F. *et al.* Dietary glycemic index, glycemic load and risk of age-related cataract extraction: a case-control study in Italy. *Eur J Nutr* **54**, 475-81 (2015).

370. Andley, U.P. Effects of alpha-crystallin on lens cell function and cataract pathology. *Current Molecular Medicine* **9**, 887-892 (2009).

371. Martinez-Lapiscina, E.H. *et al.* [Dietary fat intake and incidence of cataracts: The SUN Prospective study in the cohort of Navarra, Spain]. *Med Clin (Barc)* **134**, 194-201 (2010).

372. Lu, M. *et al.* Dietary fat intake and early age-related lens opacities. *Am J Clin Nutr* **81**, 773-9 (2005).

373. Lu, M. *et al.* Dietary linolenic acid intake is positively associated with five-year change in eye lens nuclear density. *J Am Coll Nutr* **26**, 133-40 (2007).

374. Lyle, B.J., Mares-Perlman, J.A., Klein, B.E., Klein, R. & Greger, J.L. Antioxidant intake and risk of incident age-related nuclear cataracts in the Beaver Dam Eye Study. *Am J Epidemiol* **149**, 801-9 (1999).

375. Jacques, P.F. *et al.* Long-term nutrient intake and early age-related nuclear lens opacities. *Arch Ophthalmol* **119**, 1009-19 (2001).

376. Valero, M.P., Fletcher, A.E., De Stavola, B.L., Vioque, J. & Alepuz, V.C. Vitamin C is associated with reduced risk of cataract in a Mediterranean population. *J Nutr* **132**, 1299-306 (2002).

377. Tan, A.G. *et al.* Antioxidant nutrient intake and the long-term incidence of age-related cataract: the Blue Mountains Eye Study. *Am J Clin Nutr* **87**, 1899-905 (2008).

378. Ravindran, R.D. *et al.* Inverse association of vitamin C with cataract in older people in India. *Ophthalmology* **118**, 1958-1965 e2 (2011).

379. Vitale, S. *et al.* Plasma antioxidants and risk of cortical and nuclear cataract. *Epidemiology* **4**, 195-203 (1993).

380. Ferrigno, L. *et al.* Associations between plasma levels of vitamins and cataract in the Italian-American Clinical Trial of Nutritional Supplements and Age-Related Cataract (CTNS): CTNS Report #2. *Ophthalmic Epidemiol* **12**, 71-80 (2005).

381. Robertson, J.M., Donner, A.P. & Trevithick, J.R. A possible role for vitamins C and E in cataract prevention. *Am J Clin Nutr* **53**, 346S-351S (1991).

382. Age-Related Eye Disease Study Research, G. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss: AREDS report no. 9. *Archives of Ophthalmology (Chicago, Ill.: 1960)* **119**, 1439-1452 (2001).

383. Christen, W.G. *et al.* Age-related cataract in a randomized trial of vitamins E and C in men. *Arch Ophthalmol* **128**, 1397-405 (2010).

384. Mathew, M.C., Ervin, A.M., Tao, J. & Davis, R.M. Antioxidant vitamin supplementation for preventing and slowing the progression of age-related cataract. *Cochrane Database Syst Rev*, CD004567 (2012).

385. Leske, M.C. *et al.* Antioxidant vitamins and nuclear opacities: the longitudinal study of cataract. *Ophthalmology* **105**, 831-6 (1998).

386. Jacques, P.F. *et al.* Long-term nutrient intake and 5-year change in nuclear lens opacities. *Arch Ophthalmol* **123**, 517-26 (2005).

387. Mares, J.A. *et al.* Healthy diets and the subsequent prevalence of nuclear cataract in women. *Arch Ophthalmol* **128**, 738-49 (2010).

388. Rautiainen, S., Lindblad, B.E., Morgenstern, R. & Wolk, A. Vitamin C supplements and the risk of age-related cataract: a population-based prospective cohort study in women. *Am J Clin Nutr* **91**, 487-93 (2010).

389. Zheng Selin, J., Rautiainen, S., Lindblad, B.E., Morgenstern, R. & Wolk, A. High-dose supplements of vitamins C and E, low-dose multivitamins, and the risk of age-related cataract: a population-based prospective cohort study of men. *Am J Epidemiol* **177**, 548-55 (2013).

390. Wei, L., Liang, G., Cai, C. & Lv, J. Association of vitamin C with the risk of age-

related cataract: a meta-analysis. *Acta Ophthalmol* **94**, e170-6 (2016).

391. Eaton, J.W. Is the lens canned? *Free Radic Biol Med* **11**, 207-13 (1991).

392. Hegde, K.R. & Varma, S.D. Protective effect of ascorbate against oxidative stress in the mouse lens. *Biochim Biophys Acta* **1670**, 12-8 (2004).

393. Nimse, S.B. & Pal, D. Free radicals, natural antioxidants, and their reaction mechanisms. *RSC Advances* **5**, 27986-28006 (2015).

394. Varma, S.D., Srivastava, V.K. & Richards, R.D. Photoperoxidation in lens and cataract formation: preventive role of superoxide dismutase, catalase and vitamin C. *Ophthalmic Res* **14**, 167-75 (1982).

395. Linklater, H.A., Dzialoszynski, T., McLeod, H.L., Sanford, S.E. & Trevithick, J.R. Modelling cortical cataractogenesis. XI. Vitamin C reduces gamma-crystallin leakage from lenses in diabetic rats. *Exp Eye Res* **51**, 241-7 (1990).

396. Babizhayev, M.A., Li, D.W.C., Kasus-Jacobi, A., Žorić, L. & Alió, J.L. *Studies on the Cornea and Lens*, (Springer New York, 2014).

397. Umapathy, A., Donaldson, P. & Lim, J. Antioxidant delivery pathways in the anterior eye. *Biomed Res Int* **2013**, 207250 (2013).

398. Seddon, J.M. *et al.* The use of vitamin supplements and the risk of cataract among US male physicians. *Am J Public Health* **84**, 788-92 (1994).

399. Chasan-Taber, L. *et al.* A prospective study of vitamin supplement intake and cataract extraction among U.S. women. *Epidemiology* **10**, 679-84 (1999).

400. Zhang, Y., Jiang, W., Xie, Z., Wu, W. & Zhang, D. Vitamin E and risk of age-related cataract: a meta-analysis. *Public Health Nutr* **18**, 2804-14 (2015).

401. Cui, Y.H., Jing, C.X. & Pan, H.W. Association of blood antioxidants and vitamins with risk of age-related cataract: a meta-analysis of observational studies. *Am J Clin Nutr* **98**, 778-86 (2013).

402. McNeil, J.J. *et al.* Vitamin E supplementation and cataract: randomized controlled trial. *Ophthalmology* **111**, 75-84 (2004).

403. Christen, W.G., Glynn, R.J., Chew, E.Y. & Buring, J.E. Vitamin E and age-related cataract in a randomized trial of women. *Ophthalmology* **115**, 822-829 e1 (2008).

404. Bieri, J.G., Corash, L. & Hubbard, V.S. Medical uses of vitamin E. *N Engl J Med* **308**, 1063-71 (1983).

405. Nakazawa, Y., Nagai, N., Ishimori, N., Oguchi, J. & Tamura, H. Administration of antioxidant compounds affects the lens chaperone activity and prevents the onset of cataracts. *Biomed Pharmacother* **95**, 137-143 (2017).

406. Krepler, K. & Schmid, R. Alpha-tocopherol in plasma, red blood cells and lenses with and without cataract. *Am J Ophthalmol* **139**, 266-70 (2005).

407. Mares-Perlman, J.A. *et al.* Diet and nuclear lens opacities. *Am J Epidemiol* **141**, 322-34 (1995).

408. Delcourt, C. *et al.* Risk factors for cortical, nuclear, and posterior subcapsular cataracts: the POLA study. Pathologies Oculaires Liees a l'Age. *Am J Epidemiol* **151**, 497-504 (2000).

409. Dherani, M. *et al.* Blood levels of vitamin C, carotenoids and retinol are inversely associated with cataract in a North Indian population. *Invest Ophthalmol Vis Sci* **49**, 3328-35 (2008).

410. Gale, C.R., Hall, N.F., Phillips, D.I. & Martyn, C.N. Plasma antioxidant vitamins and carotenoids and age-related cataract. *Ophthalmology* **108**, 1992-8 (2001).

411. Wang, A., Han, J., Jiang, Y. & Zhang, D. Association of vitamin A and beta-carotene with risk for age-related cataract: a meta-analysis. *Nutrition* **30**, 1113-21 (2014).

412. Glaser, T.S. *et al.* The Association of Dietary Lutein plus Zeaxanthin and B Vitamins with Cataracts in the Age-Related Eye Disease Study: AREDS Report No. 37. *Ophthalmology* **122**, 1471-9 (2015).

413. Sen, S.K., Pukazhvanthen, P. & Abraham, R. Plasma Homocysteine, Folate and Vitamin B(12) levels in senile cataract. *Indian J Clin Biochem* **23**, 255-7 (2008).

414. Sperduto, R.D. *et al.* The Linxian cataract studies. Two nutrition intervention trials. *Arch Ophthalmol* **111**, 1246-53 (1993).

415. Camacho-Barcia, M.L. *et al.* Association of Dietary Vitamin K1 Intake With the Incidence of Cataract Surgery in an Adult Mediterranean Population: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Ophthalmol* (2017).

416. Rao, P. *et al.* The Relationship Between Serum 25-Hydroxyvitamin D Levels and Nuclear Cataract in the Carotenoid Age-Related Eye Study (CAREDS), an Ancillary Study of the Women's Health Initiative. *Invest Ophthalmol Vis Sci* **56**, 4221-30 (2015).

417. Taylor, A. *et al.* Long-term intake of vitamins and carotenoids and odds of early age-related cortical and posterior subcapsular lens opacities. *Am J Clin Nutr* **75**, 540-9 (2002).

418. Vu, H.T., Robman, L., Hodge, A., McCarty, C.A. & Taylor, H.R. Lutein and zeaxanthin and the risk of cataract: the Melbourne visual impairment project. *Invest Ophthalmol Vis Sci* **47**, 3783-6 (2006).

419. Lyle, B.J. *et al.* Serum carotenoids and tocopherols and incidence of age-related nuclear cataract. *Am J Clin Nutr* **69**, 272-7 (1999).

420. Age-Related Eye Disease Study Research, G. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta

carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Archives of Ophthalmology (Chicago, Ill.: 1960)* **119**, 1417-1436 (2001).

421. Moeller, S.M. *et al.* Associations between age-related nuclear cataract and lutein and zeaxanthin in the diet and serum in the Carotenoids in the Age-Related Eye Disease Study, an Ancillary Study of the Women's Health Initiative. *Arch Ophthalmol* **126**, 354-64 (2008).

422. Karppi, J., Laukkanen, J.A. & Kurl, S. Plasma lutein and zeaxanthin and the risk of age-related nuclear cataract among the elderly Finnish population. *Br J Nutr* **108**, 148-54 (2012).

423. Ma, L. *et al.* A dose-response meta-analysis of dietary lutein and zeaxanthin intake in relation to risk of age-related cataract. *Graefes Arch Clin Exp Ophthalmol* **252**, 63-70 (2014).

424. Alves-Rodrigues, A. & Shao, A. The science behind lutein. *Toxicol Lett* **150**, 57-83 (2004).

425. Das, B.N., Thompson, J.R., Patel, R. & Rosenthal, A.R. The prevalence of age related cataract in the Asian community in Leicester: a community based study. *Eye (Lond)* **4 (Pt 5)**, 723-6 (1990).

426. Moeller, S.M. *et al.* Overall adherence to the dietary guidelines for americans is associated with reduced prevalence of early age-related nuclear lens opacities in women. *J Nutr* **134**, 1812-9 (2004).

427. Tarwadi, K.V. & Agte, V.V. Interrelationships between nutritional status, socioeconomic factors, and lifestyle in Indian cataract patients. *Nutrition* **27**, 40-5 (2011).

428. Moise, M.M., Benjamin, L.M., Doris, T.M., Dalida, K.N. & Augustin, N.O. Role of Mediterranean diet, tropical vegetables rich in antioxidants, and sunlight exposure in blindness, cataract and glaucoma among African type 2 diabetics. *Int J Ophthalmol* **5**, 231-7 (2012).

429. Nicklas, T.A., O'Neil, C.E. & Fulgoni, V.L., 3rd. Diet quality is inversely related to cardiovascular risk factors in adults. *J Nutr* **142**, 2112-8 (2012).

430. Estruch, R. *et al.* Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med* **368**, 1279-90 (2013).

431. Rodriguez-Martin, C. *et al.* The EVIDENT diet quality index is associated with cardiovascular risk and arterial stiffness in adults. *BMC Public Health* **17**, 305 (2017).

432. Li, L., Wan, X.H. & Zhao, G.H. Meta-analysis of the risk of cataract in type 2 diabetes. *BMC Ophthalmol* **14**, 94 (2014).

433. Gu, J. *et al.* Multimorbidity in the community-dwelling elderly in urban China. *Arch Gerontol Geriatr* **68**, 62-67 (2017).

434. Jacques, P.F. *et al.* Weight status, abdominal adiposity, diabetes, and early age-related lens opacities. *Am J Clin Nutr* **78**, 400-5 (2003).

435. Klein, B.E., Klein, R., Wang, Q. & Moss, S.E. Older-onset diabetes and lens opacities. The Beaver Dam Eye Study. *Ophthalmic Epidemiol* **2**, 49-55 (1995).

436. Klein, B.E., Klein, R. & Lee, K.E. Diabetes, cardiovascular disease, selected cardiovascular disease risk factors, and the 5-year incidence of age-related cataract and progression of lens opacities: the Beaver Dam Eye Study. *Am J Ophthalmol* **126**, 782-90 (1998).

437. Leske, M.C. *et al.* Diabetes, hypertension, and central obesity as cataract risk factors in a black population: The Barbados Eye Study 1 1 Members of the Barbados Eye Study Group are listed in the Appendix at the end of this article. *Ophthalmology* **106**, 35-41 (1999).

438. Machan, C.M., Hrynchak, P.K. & Irving, E.L. Age-related cataract is associated with type 2 diabetes and statin use. *Optom Vis Sci* **89**, 1165-71 (2012).

439. Olafsdottir, E., Andersson, D.K. & Stefansson, E. The prevalence of cataract in a population with and without type 2 diabetes mellitus. *Acta Ophthalmol* **90**, 334-40 (2012).

440. Tan, J.S., Wang, J.J. & Mitchell, P. Influence of diabetes and cardiovascular disease on the long-term incidence of cataract: the Blue Mountains eye study. *Ophthalmic Epidemiol* **15**, 317-27 (2008).

441. Memon, A.F. *et al.* Age-related cataract and its types in patients with and without type 2 diabetes mellitus: A Hospital-based comparative study. *J Pak Med Assoc* **66**, 1272-1276 (2016).

442. Dugmore, W.N. & Tun, K. Glucose tolerance tests in 200 patients with senile cataract. *Br J Ophthalmol* **64**, 689-92 (1980).

443. Palimeris, G., Chimonidou, E., Papakonstantinou, P. & Droutsas, G. Glycose tolerance curve, HbA1 determination and cataract. *Metab Pediatr Syst Ophthalmol* **7**, 195-200 (1983).

444. Karasik, A. *et al.* Senile cataract and glucose intolerance: the Israel Study of glucose Intolerance Obesity and Hypertension (The Israel GOH Study). *Diabetes Care* **7**, 52-6 (1984).

445. Klein, B.E.K., Klein, R. & Lee, K.E. Diabetes, cardiovascular disease, selected cardiovascular disease risk factors, and the 5-year incidence of age-related cataract and progression of lens opacities: the beaver dam eye study. *American Journal of Ophthalmology* **126**, 782-790 (1998).

446. Chang, J.R. *et al.* Risk factors associated with incident cataracts and cataract surgery in the Age-related Eye Disease Study (AREDS): AREDS report number 32.

Ophthalmology **118**, 2113-9 (2011).

447. Obrosova, I.G., Chung, S.S. & Kador, P.F. Diabetic cataracts: mechanisms and management. *Diabetes Metab Res Rev* **26**, 172-80 (2010).

448. Liu, X.F. *et al.* Nrf2 as a target for prevention of age-related and diabetic cataracts by against oxidative stress. *Aging Cell* **16**, 934-942 (2017).

449. Alberti, K.G. *et al.* Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* **120**, 1640-5 (2009).

450. Poh, S., Mohamed Abdul, R.B., Lamoureux, E.L., Wong, T.Y. & Sabanayagam, C. Metabolic syndrome and eye diseases. *Diabetes Res Clin Pract* **113**, 86-100 (2016).

451. Chopra, R., Chander, A. & Jacob, J.J. Ocular associations of metabolic syndrome. *Indian J Endocrinol Metab* **16 Suppl 1**, S6-S11 (2012).

452. Rahimi, H.R., Omidtabrizi, A., Ghayour Mobarhan, M. & Sharifi, F. Ocular Manifestations of Metabolic Syndrome. *Razavi Int J Med* **2**, e16478 %@ 2345-6434 (2014).

453. Rajappa, M., Goyal, A. & Kaur, J. Inherited metabolic disorders involving the eye: a clinico-biochemical perspective. *Eye (Lond)* **24**, 507-18 (2010).

454. Bojarskiene, F., Cerniauskiene, L.R., Paunksnis, A. & Luksiene, D.I. [Association of metabolic syndrome components with cataract]. *Medicina (Kaunas)* **42**, 115-22 (2006).

455. Paunksnis, A. *et al.* Relation between cataract and metabolic syndrome and its components. *Eur J Ophthalmol* **17**, 605-14 (2007).

456. Galeone, C. *et al.* Metabolic syndrome, its components and risk of age-related cataract extraction: a case-control study in Italy. *Ann Epidemiol* **20**, 380-4 (2010).

457. Lindblad, B.E., Hakansson, N., Philipson, B. & Wolk, A. Metabolic syndrome components in relation to risk of cataract extraction: a prospective cohort study of women. *Ophthalmology* **115**, 1687-92 (2008).

458. Ghaem Maralani, H. *et al.* Metabolic syndrome and risk of age-related cataract over time: an analysis of interval-censored data using a random-effects model. *Invest Ophthalmol Vis Sci* **54**, 641-6 (2013).

459. Sabanayagam, C. *et al.* Metabolic syndrome components and age-related cataract: the Singapore Malay eye study. *Invest Ophthalmol Vis Sci* **52**, 2397-404 (2011).

460. Park, S. & Lee, E.H. Association between metabolic syndrome and age-related cataract. *Int J Ophthalmol* **8**, 804-11 (2015).

461. Valdes, A.M., Glass, D. & Spector, T.D. Omics technologies and the study of human ageing. *Nat Rev Genet* **14**, 601-7 (2013).

462. Andrew, T. *et al.* Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-77 (2001).

463. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899-906 (2006).

464. Moayyeri, A., Hammond, C.J., Valdes, A.M. & Spector, T.D. Cohort profile: TwinsUK and healthy ageing twin study. *International journal of epidemiology*, dyr207 %U http://ije.oxfordjournals.org/content/early/2012/01/08/ije.dyr207.short (2012).

465. Moayyeri, A., Hammond, C.J., Hart, D.J. & Spector, T.D. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet* **16**, 144-9 (2013).

466. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741-6 (2007).

467. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

468. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* **44**, 955-959 (2012).

469. Consortium, U.K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).

470. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-50 (2014).

471. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet* **49**, 568-578 (2017).

472. Caporaso, J.G. *et al.* Moving pictures of the human microbiome. *Genome Biol* **12**, R50 (2011).

473. Hofman, A. *et al.* The Rotterdam Study: objectives and design update. *European Journal of Epidemiology* **22**, 819-829 %U http://link.springer.com/10.1007/s10654-007-9199-x (2007).

474. Hofman, A. *et al.* The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol* **28**, 889-926 (2013).

475. Ikram, M.A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850 (2017).

476. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol* **26**, 657-86 (2011).

477. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence

and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).

478. Age-Related Eye Disease Study Research, G. The age-related eye disease study (AREDS) system for classifying cataracts from photographs: AREDS report no. 4. *American Journal of Ophthalmology* **131**, 167-175 (2001).

479. Campbell, J.A. & Palit, C.D. Total digit dialing for a small area census by phone. Proceedings of the Survey Research Methods Section of the American Statistical Association, 549 (1988).

480. Loomis, S.J. *et al.* Exome Array Analysis of Nuclear Lens Opacity. *Ophthalmic Epidemiol* **25**, 215-219 (2018).

481. Mitchell, P., Cumming, R.G., Attebo, K. & Panchapakesan, J. Prevalence of cataract in Australia: the Blue Mountains eye study. *Ophthalmology* **104**, 581-8 (1997).

482. Murthy, G.V. *et al.* Prevalence of lens opacities in North India: the INDEYE feasibility study. *Invest Ophthalmol Vis Sci* **48**, 88-95 (2007).

483. Lavanya, R. *et al.* Methodology of the Singapore Indian Chinese Cohort (SICC) Eye Study: Quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiology* **16**, 325-336 %U http://www.tandfonline.com/doi/full/10.3109/09286580903144738 (2009).

484. Foong, A.W.P. *et al.* Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiology* **14**, 25-35 (2007).

485. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255-66 (2008).

486. Stoltenberg, S.F. Coming to terms with heritability. *Genetica* **99**, 89-96 (1997).

487. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).

488. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

489. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 %U http://www.nature.com/doifinder/10.1038/ng.3211 (2015).

490. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**, e1003348 (2013).

491. Hirschhorn, J.N. & Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**, 95-108 (2005).

492. Austin, M.A. Genetic epidemiology : methods and applications, (CABI, Wallingford, 2013).

493. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).

494. Nurnberg, S.T. *et al.* From Loci to Biology: Functional Genomics of Genome-Wide Association for Coronary Disease. *Circ Res* **118**, 586-606 (2016).

495. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).

496. Morgenthaler, S. & Thilly, W.G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* **615**, 28-56 (2007).

497. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-21 (2008).

498. Pan, W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol* **33**, 497-507 (2009).

499. Price, A.L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**, 832-8 (2010).

500. Wu, M.C. *et al.* Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet* **86**, 929-42 (2010).

501. Neale, B.M. *et al.* Testing for an unusual distribution of rare variants. *PLoS Genet* **7**, e1001322 (2011).

502. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).

503. Asimit, J.L., Day-Williams, A.G., Morris, A.P. & Zeggini, E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum Hered* **73**, 84-94 (2012).

504. Chen, L.S., Hsu, L., Gamazon, E.R., Cox, N.J. & Nicolae, D.L. An exponential combination procedure for set-based association tests in sequencing studies. *Am J Hum Genet* **91**, 977-86 (2012).

505. Derkach, A., Lawless, J.F. & Sun, L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol* **37**, 110-21 (2013).

506. Sun, J., Zheng, Y. & Hsu, L. A unified mixed-effects model for rare-variant association in sequencing studies. *Genet Epidemiol* **37**, 334-44 (2013).

507. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014).

508. Madsen, B.E. & Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**, e1000384 (2009).

509. Han, F. & Pan, W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* **70**, 42-54 (2010).

510. Lin, D.Y. & Tang, Z.Z. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* **89**, 354-67 (2011).

511. Liu, D.J. & Leal, S.M. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet* **6**, e1001156 (2010).

512. Harris, M.L., Smith, G.T. & Brown, N.A. Inter and intra observer reproducibility of the new Oxford CCD Scheimpflug camera. *Eye (Lond)* **5 (Pt 4)**, 487-90 (1991).

513. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).

514. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* **11**, 407-9 (2014).

515. Pei, Y.F., Zhang, L., Li, J. & Deng, H.W. Analyses and comparison of imputation-based association methods. *PLoS One* **5**, e10827 (2010).

516. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).

517. Liu, Q., Nicolae, D.L. & Chen, L.S. Marbled inflation from population structure in gene-based association studies with rare variants. *Genet Epidemiol* **37**, 286-92 (2013).

518. Zawistowski, M. *et al.* Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests. *Eur J Hum Genet* **22**, 1137-44 (2014).

519. Pirie, A., Wood, A., Lush, M., Tyrer, J. & Pharoah, P.D. The effect of rare variants on inflation of the test statistics in case-control analyses. *BMC Bioinformatics* **16**, 53 (2015).

520. Fuentes Fajardo, K.V. *et al.* Detecting false-positive signals in exome sequencing. *Hum Mutat* **33**, 609-13 (2012).

521. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-44 (2013).

522. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 %U http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq340 (2010).

523. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Hum Mol Genet* **24**, 7445-9 (2015).

524. Yang, J., Zeng, J., Goddard, M.E., Wray, N.R. & Visscher, P.M. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet* **49**, 1304-1310

(2017).

525. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-21 (2012).

526. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-20 (2015).

527. Krishna Kumar, S., Feldman, M.W., Rehkopf, D.H. & Tuljapurkar, S. Limitations of GCTA as a solution to the missing heritability problem. *Proc Natl Acad Sci U S A* **113**, E61-70 (2016).

528. Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-85 (2008).

529. Vejux, A., Samadi, M. & Lizard, G. Contribution of cholesterol and oxysterols in the physiopathology of cataract: implication for the development of pharmacological treatments. *J Ophthalmol* **2011**, 471947 (2011).

530. Tsaousidou, M.K. *et al.* Sequence alterations within CYP7B1 implicate defective cholesterol homeostasis in motor-neuron degeneration. *Am J Hum Genet* **82**, 510-5 (2008).

531. Gilliam, J.C. & Wensel, T.G. TRP channel gene expression in the mouse retina. *Vision Res* **51**, 2440-52 (2011).

532. Wistow, G. *et al.* Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium. *Mol Vis* **8**, 185-95 (2002).

533. Lovicu, F.J. & McAvoy, J.W. Growth factor regulation of lens development. *Dev Biol* **280**, 1-14 (2005).

534. Bennett, T.M., Mackay, D.S., Siegfried, C.J. & Shiels, A. Mutation of the melastatin-related cation channel, TRPM3, underlies inherited cataract and glaucoma. *PLoS One* **9**, e104000 (2014).

535. Heon, E. *et al.* A progressive autosomal recessive cataract locus maps to chromosome 9q13-q22. *Am J Hum Genet* **68**, 772-7 (2001).

536. Grimm, C., Kraft, R., Sauerbruch, S., Schultz, G. & Harteneck, C. Molecular and functional characterization of the melastatin-related cation channel TRPM3. *J Biol Chem* **278**, 21493-501 (2003).

537. Shaham, O. *et al.* Pax6 regulates gene expression in the vertebrate lens through miR-204. *PLoS Genet* **9**, e1003357 (2013).

538. Yang, Y. *et al.* TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* **40**, 1016-23 (2010).

539. Lachke, S.A. *et al.* Mutations in the RNA granule component TDRD7 cause

cataract and glaucoma. *Science* **331**, 1571-6 (2011).

540. Thorleifsson, G. *et al.* Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma. *Nat Genet* **42**, 906-9 (2010).

541. Biswas, S.K., Brako, L. & Lo, W.K. Massive formation of square array junctions dramatically alters cell shape but does not cause lens opacity in the cav1-KO mice. *Exp Eye Res* **125**, 9-19 (2014).

542. Garg, A. *et al.* Whole exome sequencing identifies de novo heterozygous CAV1 mutations associated with a novel neonatal onset lipodystrophy syndrome. *Am J Med Genet A* **167A**, 1796-806 (2015).

543. Castellini, M. *et al.* Palm is expressed in both developing and adult mouse lens and retina. *BMC Ophthalmol* **5**, 14 (2005).

544. Gold, M.G. *et al.* AKAP2 anchors PKA with aquaporin-0 to support ocular lens transparency. *EMBO Mol Med* **4**, 15-26 (2012).

545. Jia, Y. *et al.* MMP-2, MMP-3, TIMP-1, TIMP-2, and TIMP-3 protein levels in human aqueous humor: relationship with axial length. *Invest Ophthalmol Vis Sci* **55**, 3922-8 (2014).

546. Ashworth Briggs, E.L., Toh, T., Eri, R., Hewitt, A.W. & Cook, A.L. TIMP1, TIMP2, and TIMP4 are increased in aqueous humor from primary open angle glaucoma patients. *Mol Vis* **21**, 1162-72 (2015).

547. West-Mays, J.A. & Pino, G. Matrix Metalloproteinases as Mediators of Primary and Secondary Cataracts. *Expert Rev Ophthalmol* **2**, 931-938 (2007).

548. Jiang, D. & McPeek, M.S. Robust rare variant association testing for quantitative traits in samples with related individuals. *Genet Epidemiol* **38**, 10-20 (2014).

549. Evangelou, E. & Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* **14**, 379-89 (2013).

550. Zeggini, E. & Ioannidis, J.P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191-201 (2009).

551. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* **88**, 586-98 (2011).

552. Begum, F., Ghosh, D., Tseng, G.C. & Feingold, E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res* **40**, 3777-84 (2012).

553. Manchia, M. *et al.* The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One* **8**, e76295 (2013).

554. Ioannidis, J.P., Patsopoulos, N.A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e841 (2007).

555. Ioannidis, J.P. *et al.* Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol* **37**, 120-32 (2008).

556. Ioannidis, J.P. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* **64**, 203-13 (2007).

557. Neupane, B., Loeb, M., Anand, S.S. & Beyene, J. Meta-analysis of genetic association studies under heterogeneity. *Eur J Hum Genet* **20**, 1174-81 (2012).

558. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).

559. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557-60 (2003).

560. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497-508 (2014).

561. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).

562. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-22 (2011).

563. Spain, S.L. & Barrett, J.C. Strategies for fine-mapping complex traits. *Hum Mol Genet* **24**, R111-9 (2015).

564. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807-812 %U http://www.nature.com/doifinder/10.1038/ejhg.2011.39 (2011).

565. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American Journal of Human Genetics* **88**, 586-598 (2011).

566. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genetic Epidemiology* **35**, 809-822 %U http://doi.wiley.com/10.1002/gepi.20630 (2011).

567. Li, M.-X., Gui, H.-S., Kwan, J.S.H. & Sham, P.C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *American Journal of Human Genetics* **88**, 283-293 (2011).

568. Segrè, A.V. *et al.* Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS Genetics* **6**, e1001058 %U http://dx.plos.org/10.1371/journal.pgen.1001058 (2010).

569. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.

Nucleic Acids Res **40**, D930-4 (2012).

570. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084-1089 %U http://www.nature.com/doifinder/10.1038/ng.2394 (2012).

571. Buil, A. *et al.* Quantifying the degree of sharing of genetic and non-genetic causes of gene expression variability across four tissues. %U http://biorxiv.org/lookup/doi/10.1101/053355. (2016).

572. Lachke, S.A. *et al.* iSyTE : I ntegrated Sy stems T ool for E ye Gene Discovery. *Investigative Opthalmology & Visual Science* **53**, 1617 %U http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.11-8839 (2012).

573. Fahrenbach, J.P., Andrade, J. & McNally, E.M. The CO-Regulation Database (CORD): a tool to identify coordinately expressed genes. *PLoS One* **9**, e90408 (2014).

574. Liao, J. *et al.* Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract. *Hum Mol Genet* **23**, 6119-28 (2014).

575. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).

576. Gates, L.A. *et al.* Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J Biol Chem* **292**, 14456-14472 (2017).

577. Gupta, D., Harvey, S.A., Kenchegowda, D., Swamynathan, S. & Swamynathan, S.K. Regulation of mouse lens maturation and gene expression by Kruppel-like factor 4. *Exp Eye Res* **116**, 205-18 (2013).

578. Dateki, S. *et al.* OTX2 mutation in a patient with anophthalmia, short stature, and partial growth hormone deficiency: functional studies using the IRBP, HESX1, and POU1F1 promoters. *J Clin Endocrinol Metab* **93**, 3697-702 (2008).

579. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev* **30**, 881-91 (2016).

580. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660-665 %U http://www.sciencemag.org/cgi/doi/10.1126/science.aaa0355 (2015).

581. Aguet, F. *et al.* Local genetic effects on gene expression across 44 human tissues %U http://biorxiv.org/content/early/2016/09/09/074450.abstract. *bioRxiv* (2016).

582. Amaral, P.P. *et al.* Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. *RNA* **15**, 2013-2027 %U http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.1705309 (2009).

583. Zhang, X., Wang, L., Wang, J., Dong, B. & Li, Y. Coralliform cataract caused by

a novel connexin46 (GJA3) mutation in a Chinese family. *Molecular Vision* **18**, 203-210 (2012).

584. Beyer, E.C., Ebihara, L. & Berthoud, V.M. Connexin mutants and cataracts. *Frontiers in Pharmacology* **4**, 43 (2013).

585. Hejtmancik, J.F. *et al.* Lens Biology and Biochemistry. in *Progress in Molecular Biology and Translational Science*, Vol. 134 169-201 %@ 978-0-12-801059-4 %U http://linkinghub.elsevier.com/retrieve/pii/S187711731500068X (Elsevier, 2015).

586. Boelens, W.C. Cell biological roles of αB-crystallin. *Progress in Biophysics and Molecular Biology* **115**, 3-10 (2014).

587. Christopher, K.L. *et al.* Alpha-crystallin-mediated protection of lens cells against heat and oxidative stress-induced cell death. *Biochimica Et Biophysica Acta* **1843**, 309-315 (2014).

588. Pras, E. *et al.* A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. *Investigative Ophthalmology & Visual Science* **41**, 3511-3515 (2000).

589. Sacconi, S. *et al.* A novel CRYAB mutation resulting in multisystemic disease. *Neuromuscular disorders: NMD* **22**, 66-72 (2012).

590. Zhou, P., Luo, Y., Liu, X., Fan, L. & Lu, Y. Down-regulation and CpG island hypermethylation of CRYAA in age-related nuclear cataract. *The FASEB Journal* **26**, 4897-4902 %U http://www.fasebj.org/cgi/doi/10.1096/fj.12-213702 (2012).

591. Zhou, H.Y. *et al.* Quantitative proteomics analysis by iTRAQ in human nuclear cataracts of different ages and normal lens nuclei. *Proteomics. Clinical Applications* **9**, 776-786 (2015).

592. Fantes, J. *et al.* Mutations in SOX2 cause anophthalmia. *Nature Genetics* **33**, 461-463 (2003).

593. Kondoh, H., Uchikawa, M. & Kamachi, Y. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *The International Journal of Developmental Biology* **48**, 819-827 (2004).

594. Donner, A.L., Episkopou, V. & Maas, R.L. Sox2 and Pou2f1 interact to control lens and olfactory placode development. *Developmental Biology* **303**, 784-799 (2007).

595. Liu, W., Lagutin, O.V., Mende, M., Streit, A. & Oliver, G. Six3 activation of Pax6 expression is essential for mammalian lens induction and specification. *The EMBO journal* **25**, 5383-5395 (2006).

596. Shimada, N., Aya-Murata, T., Reza, H.M. & Yasuda, K. Cooperative action between L-Maf and Sox2 on δ-crystallin gene expression during chick lens development. *Mechanisms of Development* **120**, 455-465 %U http://linkinghub.elsevier.com/retrieve/pii/S0925477303000029 (2003).

597. Manthey, A.L. *et al.* Loss of Sip1 leads to migration defects and retention of ectodermal markers during lens development. *Mech Dev* **131**, 86-110 (2014).

598. Maddala, R., Nagendran, T., Lang, R.A., Morozov, A. & Rao, P.V. Rap1 GTPase is required for mouse lens epithelial maintenance and morphogenesis. *Dev Biol* **406**, 74-91 (2015).

599. Wride, M.A., Geatrell, J. & Guggenheim, J.A. Proteases in eye development and disease. *Birth Defects Research Part C: Embryo Today: Reviews* **78**, 90-105 %U http://doi.wiley.com/10.1002/bdrc.20063 (2006).

600. Liu, K. *et al.* Altered ubiquitin causes perturbed calcium homeostasis, hyperactivation of calpain, dysregulated differentiation, and cataract. *Proceedings of the National Academy of Sciences* **112**, 1071-1076 %U http://www.pnas.org/lookup/doi/10.1073/pnas.1404059112 (2015).

601. de Bie, P., van de Sluis, B., Klomp, L. & Wijmenga, C. The many faces of the copper metabolism protein MURR1/COMMD1. *The Journal of Heredity* **96**, 803-811 (2005).

602. Mao, X. *et al.* COMMD1 (copper metabolism MURR1 domain-containing protein 1) regulates Cullin RING ligases by preventing CAND1 (Cullin-associated Nedd8-dissociated protein 1) binding. *The Journal of Biological Chemistry* **286**, 32355-32365 (2011).

603. Guilherme, A., Soriano, N.A., Furcinitti, P.S. & Czech, M.P. Role of EHD1 and EHBP1 in perinuclear sorting and insulin-regulated GLUT4 recycling in 3T3-L1 adipocytes. *J Biol Chem* **279**, 40062-75 (2004).

604. Arya, P. *et al.* The endocytic recycling regulatory protein EHD1 Is required for ocular lens development. *Dev Biol* **408**, 41-55 (2015).

605. Li, Z. *et al.* A novel Rab10-EHBP1-EHD2 complex essential for the autophagic engulfment of lipid droplets. *Sci Adv* **2**, e1601470 (2016).

606. Wang, X. *et al.* Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Human Molecular Genetics* **22**, 2303-2311 %U https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddt064 (2013).

607. Mila-Villarroel, R. *et al.* Comparison and evaluation of the reliability of indexes of adherence to the Mediterranean diet. *Public Health Nutr* **14**, 2338-45 (2011).

608. Asghari, G., Mirmiran, P., Yuzbashian, E. & Azizi, F. A systematic review of diet quality indices in relation to obesity. *Br J Nutr*, 1-11 (2017).

609. Dinu, M., Pagliai, G., Casini, A. & Sofi, F. Mediterranean diet and multiple health outcomes: an umbrella review of meta-analyses of observational studies and randomised trials. *Eur J Clin Nutr* (2017).

610. Galor, A., Gardener, H., Pouyeh, B., Feuer, W. & Florez, H. Effect of a

Mediterranean dietary pattern and vitamin D levels on Dry Eye syndrome. *Cornea* **33**, 437-41 (2014).

611. Merle, B.M., Silver, R.E., Rosner, B. & Seddon, J.M. Adherence to a Mediterranean diet, genetic susceptibility, and progression to advanced macular degeneration: a prospective cohort study. *Am J Clin Nutr* **102**, 1196-206 (2015).

612. Hogg, R.E. *et al.* Mediterranean Diet Score and Its Association with Age-Related Macular Degeneration: The European Eye Study. *Ophthalmology* **124**, 82-89 (2017).

613. Kennedy, E. Putting the pyramid into action: the Healthy Eating Index and Food Quality Score. *Asia Pac J Clin Nutr* **17 Suppl 1**, 70-4 (2008).

614. Nicholson, J.K. & Lindon, J.C. Systems biology: Metabonomics. *Nature* **455**, 1054-6 (2008).

615. Menni, C., Zierer, J., Valdes, A.M. & Spector, T.D. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nat Rev Rheumatol* **13**, 174-181 (2017).

616. Tektonidou, M.G. & Ward, M.M. Validity of clinical associations of biomarkers in translational research studies: the case of systemic autoimmune diseases. *Arthritis Res Ther* **12**, R179 (2010).

617. Kapoor, S.R. *et al.* Metabolic profiling predicts response to anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis. *Arthritis Rheum* **65**, 1448-56 (2013).

618. Patti, G.J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-9 (2012).

619. Marion, D. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics* **12**, 3006-25 (2013).

620. Fan, T.W. & Lane, A.N. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc* **92-93**, 18-53 (2016).

621. Lei, Z., Huhman, D.V. & Sumner, L.W. Mass spectrometry strategies in metabolomics. *J Biol Chem* **286**, 25435-42 (2011).

622. Sellick, C.A. *et al.* Evaluation of extraction processes for intracellular metabolite profiling of mammalian cells: matching extraction approaches to cell type and metabolite targets. *Metabolomics* **6**, 427-438 (2010).

623. Tan, S.Z., Begley, P., Mullard, G., Hollywood, K.A. & Bishop, P.N. Introduction to metabolomics and its applications in ophthalmology. *Eye (Lond)* **30**, 773-83 (2016).

624. Kryczka, T., Wylegala, E., Dobrowolski, D. & Midelfart, A. NMR spectroscopy of human eye tissues: a new insight into ocular biochemistry. *ScientificWorldJournal* **2014**, 546192 (2014).

625. Snytnikova, O.A. *et al.* Metabolomics of the human aqueous humor. *Metabolomics* **13**, 5 (2016).

626. Barba, I. *et al.* Metabolic fingerprints of proliferative diabetic retinopathy: an 1H-NMR-based metabonomic approach using vitreous humor. *Invest Ophthalmol Vis Sci* **51**, 4416-21 (2010).

627. Osborn, M.P. *et al.* Metabolome-wide association study of neovascular age-related macular degeneration. *PLoS One* **8**, e72737 (2013).

628. Burgess, L.G. *et al.* Metabolome-Wide Association Study of Primary Open Angle Glaucoma. *Invest Ophthalmol Vis Sci* **56**, 5020-8 (2015).

629. Panduro, A., Rivera-Iniguez, I., Sepulveda-Villegas, M. & Roman, S. Genes, emotions and gut microbiota: The next frontier for the gastroenterologist. *World J Gastroenterol* **23**, 3030-3042 (2017).

630. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* **32**, 834-41 (2014).

631. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220-30 (2012).

632. Ranjan, R., Rani, A., Metwally, A., McGee, H.S. & Perkins, D.L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun* **469**, 967-77 (2016).

633. Blaut, M. & Clavel, T. Metabolic diversity of the intestinal microbiota: implications for health and disease. *J Nutr* **137**, 751S-5S (2007).

634. Rios-Covian, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front Microbiol* **7**, 185 (2016).

635. Lupton, J.R. Microbial degradation products influence colon cancer risk: the butyrate controversy. *J Nutr* **134**, 479-82 (2004).

636. Ley, R.E., Turnbaugh, P.J., Klein, S. & Gordon, J.I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022-3 (2006).

637. Frank, D.N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**, 13780-5 (2007).

638. Dicksved, J. *et al.* Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J* **2**, 716-27 (2008).

639. Turnbaugh, P.J., Backhed, F., Fulton, L. & Gordon, J.I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213-23 (2008).

640. Gonzalez, A. *et al.* The mind-body-microbial continuum. *Dialogues Clin*

Neurosci **13**, 55-62 (2011).

641. Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L. & Gordon, J.I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327-36 (2011).

642. Zinkernagel, M.S. *et al.* Association of the Intestinal Microbiome with the Development of Neovascular Age-Related Macular Degeneration. *Sci Rep* **7**, 40826 (2017).

643. Caldwell, R.B., Zhang, W., Romero, M.J. & Caldwell, R.W. Vascular dysfunction in retinopathy-an emerging role for arginase. *Brain Res Bull* **81**, 303-9 (2010).

644. Ishikawa, M. Abnormalities in glutamate metabolism and excitotoxicity in the retinal diseases. *Scientifica (Cairo)* **2013**, 528940 (2013).

645. Liu, A., Lin, Y., Terry, R., Nelson, K. & Bernstein, P.S. Role of long-chain and very-long-chain polyunsaturated fatty acids in macular degenerations and dystrophies. *Clin Lipidol* **6**, 593-613 (2011).

646. Gorusupudi, A., Liu, A., Hageman, G.S. & Bernstein, P.S. Associations of human retinal very long-chain polyunsaturated fatty acids with dietary lipid biomarkers. *J Lipid Res* **57**, 499-508 (2016).

647. Turnbaugh, P.J. *et al.* The human microbiome project. *Nature* **449**, 804-10 (2007).

648. Costello, E.K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-7 (2009).

649. Turnbaugh, P.J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480-4 (2009).

650. Wu, G.D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-8 (2011).

651. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222-7 (2012).

652. Matijasic, B.B. *et al.* Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur J Nutr* **53**, 1051-64 (2014).

653. David, L.A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-63 (2014).

654. O'Keefe, S.J. *et al.* Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun* **6**, 6342 (2015).

655. Singh, R.K. *et al.* Influence of diet on the gut microbiome and implications for human health. *J Transl Med* **15**, 73 (2017).

656. Yonova-Doing, E. *et al.* Genetic and Dietary Factors Influencing the Progression of Nuclear Cataract. *Ophthalmology* **123**, 1237-44 (2016).

657. Bingham, S.A. *et al.* Nutritional methods in the European Prospective

Investigation of Cancer in Norfolk. *Public Health Nutr* **4**, 847-58 (2001).

658. Teucher, B. *et al.* Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet* **10**, 734-48 (2007).

659. McCance, R.A., Widdowson, E.M., Holland, B., Welch, A.A. & Buss, D.H. McCance and Widdowson's the composition of foods. 2nd suppl, Vegetable dishes, (Royal Society of Chemistry, 1991).

660. Willett, W. & Stampfer, M.J. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* **124**, 17-27 (1986).

661. McCance, R.A.C.o.f. McCance and Widdowson's the composition of foods, (2015).

662. Fung, T.T. *et al.* Mediterranean diet and incidence of and mortality from coronary heart disease and stroke in women. *Circulation* **119**, 1093-100 (2009).

663. Guenther, P.M. *et al.* Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet* **113**, 569-80 (2013).

664. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-7 (2005).

665. MacKinnon, D.P., Fairchild, A.J. & Fritz, M.S. Mediation analysis. *Annu Rev Psychol* **58**, 593-614 (2007).

666. Gottlieb, D.J. *et al.* Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med* **164**, 1655-9 (2001).

667. Christensen, K., Gaist, D., Vaupel, J.W. & McGue, M. Genetic contribution to rate of change in functional abilities among Danish twins aged 75 years or more. *Am J Epidemiol* **155**, 132-9 (2002).

668. Steves, C.J., Jackson, S.H. & Spector, T.D. Cognitive change in older women using a computerised battery: a longitudinal quantitative genetic twin study. *Behav Genet* **43**, 468-79 (2013).

669. Kanthan, G.L., Mitchell, P., Burlutsky, G. & Wang, J.J. Fasting blood glucose levels and the long-term incidence and progression of cataract -- the Blue Mountains Eye Study. *Acta Ophthalmol* **89**, e434-8 (2011).

670. Shui, Y.B. *et al.* The gel state of the vitreous and ascorbate-dependent oxygen consumption: relationship to the etiology of nuclear cataracts. *Arch Ophthalmol* **127**, 475-82 (2009).

671. Beebe, D.C., Holekamp, N.M. & Shui, Y.B. Oxidative damage and the prevention of age-related cataracts. *Ophthalmic Res* **44**, 155-65 (2010).

672. Kisic, B., Miric, D., Zoric, L., Ilic, A. & Dragojevic, I. Antioxidant capacity of lenses with age-related cataract. *Oxid Med Cell Longev* **2012**, 467130 (2012).

673. Racz, P. & Ordogh, M. Investigations on trace elements in normal and senile

cataractous lenses. Activation analysis of copper, zinc, manganese, cobalt, rubidium, scandium, and nickel. *Albrecht Von Graefes Arch Klin Exp Ophthalmol* **204**, 67-72 (1977).

674. Cekic, O. *et al.* Nickel, chromium, manganese, iron and aluminum levels in human cataractous and normal lenses. *Ophthalmic Res* **31**, 332-6 (1999).

675. Reddy, V.N., Kasahara, E., Hiraoka, M., Lin, L.R. & Ho, Y.S. Effects of variation in superoxide dismutases (SOD) on oxidative stress and apoptosis in lens epithelium. *Exp Eye Res* **79**, 859-68 (2004).

676. Tarwadi, K. & Agte, V. Linkages of antioxidant, micronutrient, and socioeconomic status with the degree of oxidative stress and lens opacity in indian cataract patients. *Nutrition* **20**, 261-7 (2004).

677. Kuzniarz, M., Mitchell, P., Cumming, R.G. & Flood, V.M. Use of vitamin supplements and cataract: the Blue Mountains Eye Study. *Am J Ophthalmol* **132**, 19-26 (2001).

678. Newby, P.K., Weismayer, C., Akesson, A., Tucker, K.L. & Wolk, A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. *J Nutr* **136**, 626-33 (2006).

679. Jankovic, N. *et al.* Stability of dietary patterns assessed with reduced rank regression; the Zutphen Elderly Study. *Nutr J* **13**, 30 (2014).

680. Pallister, T., Spector, T.D. & Menni, C. Twin studies advance the understanding of gene-environment interplay in human nutrigenomics. *Nutr Res Rev* **27**, 242-51 (2014).

681. Heiman, M.L. & Greenway, F.L. A healthy gastrointestinal microbiome is dependent on dietary diversity. *Mol Metab* **5**, 317-20 (2016).

682. Boskou, D.e. Olive Oil : Chemistry and Technology.

683. Stone, N.J. Fish Consumption, Fish Oil, Lipids, and Coronary Heart Disease. *Circulation* **94**, 2337 (1996).

684. Yamanouchi, T. *et al.* Clinical usefulness of serum 1,5-anhydroglucitol in monitoring glycaemic control. *Lancet* **347**, 1514-8 (1996).

685. Dungan, K.M. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. *Expert Rev Mol Diagn* **8**, 9-19 (2008).

686. Ortwerth, B.J. & Olesen, P.R. Glutathione inhibits the glycation and crosslinking of lens proteins by ascorbic acid. *Exp Eye Res* **47**, 737-50 (1988).

687. Pallister, T. *et al.* Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One* **11**, e0158568 (2016).

688. Dahl, A. *et al.* A multiple-phenotype imputation method for genetic studies. *Nat Genet* **48**, 466-72 (2016).

689. Mesa, R., Tyagi, M., Harocopos, G., Vollman, D. & Bassnett, S. Somatic Variants in the Human Lens Epithelium: A Preliminary Assessment. *Invest Ophthalmol Vis Sci* **57**, 4063-75 (2016).

690. Gasser, T. Usefulness of Genetic Testing in PD and PD Trials: A Balanced Review. *J Parkinsons Dis* **5**, 209-15 (2015).

691. Karch, C.M. & Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry* **77**, 43-51 (2015).

692. Shiels, A. & Hejtmancik, J.F. Mutations and mechanisms in congenital and age-related cataracts. *Exp Eye Res* **156**, 95-102 (2017).

693. Diehn, J.J., Diehn, M., Marmor, M.F. & Brown, P.O. Differential gene expression in anatomical compartments of the human eye. *Genome Biol* **6**, R74 (2005).

694. Datiles, M.B., 3rd *et al.* Longitudinal Study of Age-Related Cataract Using Dynamic Light Scattering: Loss of alpha-Crystallin Leads to Nuclear Cataract Development. *Ophthalmology* **123**, 248-54 (2016).

695. Oka, M., Kudo, H., Sugama, N., Asami, Y. & Takehana, M. The function of filensin and phakinin in lens transparency. *Mol Vis* **14**, 815-22 (2008).

696. Maher, G.J., Black, G.C. & Manson, F.D. Focus on molecules: lens intrinsic membrane protein (LIM2/MP20). *Exp Eye Res* **103**, 115-6 (2012).

697. Wang, T., Li, L. & Hong, W. SNARE proteins in membrane trafficking. *Traffic* (2017).

698. Yoshimoto, A., Saigou, Y., Higashi, Y. & Kondoh, H. Regulation of ocular lens development by Smad-interacting protein 1involving Foxe3 activation. *Development* **132**, 4437-4448 %U http://dev.biologists.org/cgi/doi/10.1242/dev.02022 (2005).

699. Arnold, K. *et al.* Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9**, 317-29 (2011).

700. Truscott, R.J. Age-related nuclear cataract-oxidation is the key. *Exp Eye Res* **80**, 709-25 (2005).

701. Bhuyan, K.C., Bhuyan, D.K. & Podos, S.M. Lipid peroxidation in cataract of the human. *Life Sci* **38**, 1463-71 (1986).

702. Kamei, A. Characterization of water-insoluble proteins in normal and cataractous human lens. *Jpn J Ophthalmol* **34**, 216-24 (1990).

703. Garner, W.H., Garner, M.H. & Spector, A. H2O2-induced uncoupling of bovine lens Na+,K+-ATPase. *Proc Natl Acad Sci U S A* **80**, 2044-8 (1983).

704. Lin, H. *et al.* Lens regeneration using endogenous stem cells with gain of visual function. *Nature* **531**, 323-8 (2016).

705. Fu, Q. *et al.* Generation of Functional Lentoid Bodies From Human Induced Pluripotent Stem Cells Derived From Urinary Cells. *Invest Ophthalmol Vis Sci* **58**,

517-527 (2017).

706. Pallister, T. *et al.* Food Preference Patterns in a UK Twin Cohort. *Twin Res Hum Genet* **18**, 793-805 (2015).

707. Yousri, N.A. *et al.* Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* **10**, 1005-1017 (2014).

708. Oram, R.A. *et al.* A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care* **39**, 337-44 (2016).

709. Thomas, N.J. *et al.* Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol* **6**, 122-129 (2018).

710. Hunter, D.J. Gene-environment interactions in human diseases. *Nat Rev Genet* **6**, 287-98 (2005).

711. Cordell, H.J. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* **10**, 392-404 (2009).

712. Aschard, H. *et al.* Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Hum Genet* **131**, 1591-613 (2012).

713. Makley, L.N. *et al.* Pharmacological chaperone for alpha-crystallin partially restores transparency in cataract models. *Science* **350**, 674-7 (2015).

714. Zhao, L. *et al.* Lanosterol reverses protein aggregation in cataracts. *Nature* **523**, 607-11 (2015).

715. Shanmugam, P.M. *et al.* Effect of lanosterol on human cataract nucleus. *Indian J Ophthalmol* **63**, 888-90 (2015).

716. Bai, J., Yang, F., Dong, L. & Zheng, Y. Ghrelin Protects Human Lens Epithelial Cells against Oxidative Stress-Induced Damage. *Oxid Med Cell Longev* **2017**, 1910450 (2017).

717. Gregson, J.M. *et al.* Genetic invalidation of Lp-PLA2 as a therapeutic target: Large-scale study of five functional Lp-PLA2-lowering alleles. *Eur J Prev Cardiol* **24**, 492-504 (2017).

718. Mark A. Babizhayev, D.W.-C.L., Anne Kasus-Jacobi, Lepša Žorić, Jorge L. Alió. *Studies on the Cornea and Lens*, (2015).

719. Javadzadeh, A. *et al.* Preventive effect of onion juice on selenite-induced experimental cataract. *Indian J Ophthalmol* **57**, 185-9 (2009).

720. Devi, V.G., Rooban, B.N., Sasikala, V., Sahasranamam, V. & Abraham, A. Isorhamnetin-3-glucoside alleviates oxidative stress and opacification in selenite cataract in vitro. *Toxicol In Vitro* **24**, 1662-9 (2010).

721. Manikandan, R., Thiagarajan, R., Beulaja, S., Sudhandiran, G. & Arumugam, M.

Effect of curcumin on selenite-induced cataractogenesis in Wistar rat pups. *Curr Eye Res* **35**, 122-9 (2010).

722. Sasikala, V., Rooban, B.N., Priya, S.G., Sahasranamam, V. & Abraham, A. Moringa oleifera prevents selenite-induced cataractogenesis in rat pups. *J Ocul Pharmacol Ther* **26**, 441-7 (2010).

723. Rooban, B.N., Sasikala, V., Gayathri Devi, V., Sahasranamam, V. & Abraham, A. Prevention of selenite induced oxidative stress and cataractogenesis by luteolin isolated from Vitex negundo. *Chem Biol Interact* **196**, 30-8 (2012).

724. Doganay, S. *et al.* The effect of apricots on the experimental cataract model formed by sodium selenite. *Food Chem Toxicol* **55**, 371-7 (2013).

Appendices

Appendix I: Eye examination protocols

VISIONIX VX120

3.8.7. VISIONIX measurements

A. Starting and turning off the VISIONIX

1. Switch on the socket switches for the VISIONIX and for the table on which the VISIONIX is placed
2. Switch on VISIONIX by switching on the on button on the touch screen
3. In the end of the day, press the shut down button on the Home screen
4. When the VISIONIX is off, switch off the socket switches for the VISIONIX and for the table

B. Inputting volunteer data

1. Tap the Add Patient button on the Home Screen (left icon on 9 o'clock)
2. Fill in the information fields
3. If the volunteer is already sitting in front of the camera press Scan & Diagnose, otherwise press save
4. If the next person whose data you are inputting has the same family name press Save & New

C. Finding volunteers' records

1. Tap the Patient screen button on the Home Screen (right icon on 3 o'clock)
2. Type Surname or Name in the Search Field above
3. Select the patient name on the left-hand side by taping on it and press New Diagnostics
4. If patient is not present add patient by pressing the add patient button (first button on the left in the bottom panel)

D. Editing volunteers' information

1. To edit information, select the patient name and tap Modify patient (second button on the left in the bottom panel)

2. After making modification press save
3. If you have taken measurement on the wrong name:
    a. tap the measurement on the right hand side of the screen
    b. press cut
    c. go to the right name
    d. press paste

E. Positioning the volunteer

1. Explain to the volunteer what you are going to do
2. Have the volunteer sit comfortably
3. Adjust the table high to the sitting high of the volunteer so that the camera is on the level of the volunteer's face and his/her eyes are on the level of mark lines on the head rest
4. Ask them to put their chin on the chin rest and to firmly press their forehead to the head rest
5. Ask the volunteer to look at the balloon and to try and keep their eyes wide open

F. Taking measurements

1. Select the name of the patient by taping on it on the left hand side of the screen
2. Tap New Diagnostics
3. Make sure the name of the patient is written on top of the diagnostic screen
4. Make sure that the eye of the volunteer is in the middle of the screen. To centre tap on top of the eye on the screen
5. Make sure that the option ALL is selected from the menu of available diagnostics (right top cornet of the screen)
6. Press Go
7. When the machine starts moving down, tell the volunteer that they will fell 3 puffs of air
8. If you cannot take a measurement or the machines struggles to measure press skip. Always do your best to obtain a measurement

9. If the tonometer struggles to focus on the eye do the following:
    a. Make sure the volunteer is positioned correctly
    b. If the target is not aligned with the black circle in the middle tap on the target till it aligns
    c. If the target is fuzzy and out of focus, zoom in by pressing the plus sign on the zoom on the upper right-hand side of the screen
    d. If the 4 lines of the target look double, zoom out by pressing the minus sign on the upper right-hand side of the screen
10. When the measurements are finished press Continue 2x to access the Results Summary screen
11. Advise the volunteer about their refraction, cornea and pressure readings
12. Press exit to save the results

OPTOVUE

3.8.6. OCT and Fundus measurements

NB. Imaging order: all OCT imagines for both eyes are taken first and then move to taking Fundus images

NB. All images are taken in with the lights in the room switched off

NB. You need to take 5 scans each eye with the OCT (retina map, retina 3D, ONH, disk 3D, GCC) and 2 photos each eye with the fundus camera (macula, ONH)

NB. If in doubt concerning eye pathology and always before referring volunteers to a specialist ask an ophthalmologist

A. Starting and turning off the camera and computer

1. Make sure that the main black switch on the back of the control box is switched on
2. Switch on the computer by switching on the socket switch on the wall
3. Switch on the iView camera by switching on the socket switch on the wall

4. Switch on the iCam by pressing the button on the iCam camera
5. Wait 30 seconds
6. Switch on the computer screen
7. Wait for the operating system to initialise
8. Press the iView and iCam icons to start the software
9. When finished for the day:
    - Close the iView and iCam programs
    - Select the shut down option from the main computer menu
    - When the screen switches off, switch off the socket switch on the wall

B. Inputting volunteer data: can be done on either of the programs, as the two programs share one database, you need to input the data once only

1. Go to Patient Menu
2. Press Add Patient button
3. Fill out the information fields
4. If the volunteer is already sat in front of the camera and you see the OCT menu press Scan otherwise press Save and press the iVew tab (on the left) to go to the OCT menu

C. Finding volunteers' records

1. Go to Patient Menu
2. Look for the name of the volunteer in the Patient list
3. OR input volunteer name in the search field and press Search

D. Editing volunteers' information

1. To edit information, select the patient name and click the Edit button
2. To edit visit information, select a visit date and click the Edit button

3. To add additional comments such as reasons why taking photos was difficult or unsuccessful select the patient name, click the Edit button and fill in the comments box

E. Deleting data – **NB data once deleted can not be retrieved back**

1. For OCT data – review the scan in the review window, any scan with low quality due to blinking or other factors should be deleted so that there are only 10 scans per person. For that right-click on the scan you want to discard and press delete

2. For the Fundus data – look at the photos in the review window, any photos with low quality due to blinking or other factors should be deleted so that there are only 4 photos per person. For that remember the position in the taking order of the photo you want to discard and go back to the main iCam menu. Right-click on the photo you want to discard and press delete

F. Positioning the volunteer

1. Explain to the volunteer what you are going to do

2. Have the volunteer sat comfortably

3. Adjust the table high to the sitting high of the volunteer so that the camera is on the level of the volunteer's face

4. Ask them to put their chin on the chin rest and to firmly press their forehead to the head rest

5. Their eyes should be on the level of the black lines which are visible on both sides of the head rest

6. The chin rest should be adjusted to be very slightly lower so that the volunteer leans forwards

7. Tell the volunteer that they can blink until you tell them to stare

G. Taking OCT measurements

1. After clicking Scan, the scan screen will appear

2. Move the camera close to the eye you want to measure and use the joystick to focus
3. Perform all measurements on the right eye first and then move to the left eye
4. Press Step 1
5. Select an eye (Right/OD for the right eye or Left/OS for the left eye) - the eye selected will be shown below the video screen
6. **Selecting Retina**
7. **Press Retina Map**
8. Ask the volunteer to look at the **"X"** in the centre of the target which should be straight-ahead. This is true for all retinal images

9. Press Auto-Adjust button to enhance the OCT image
10. The image should be in-between the red lines on the measurement screen
11. To adjust the image position use the mouse scroll
12. When ready to take the image ask the volunteer to blink one last time and then repeat three times "Stare"
13. After the last "Stare" press the capture button on the screen or the black button on top of the joystick
14. If you are happy with the quality of image (the bar on the right sight of the image is green and the percentage given is high >60%) go back to Step 1 for another measurement, if not press Scan Again

15. **Press retina 3D (done for everyone on their first measurement with the machine) (not needed for follow up)**
16. Make sure that the image is under the higher red line
17. Ask the volunteer to blink one last time and then repeat three times “Stare”
18. After the last “Stare” press the capture button on the screen or the black button on top of the joystick
19. Skip the Cross Line measurement
20. **Press Nerve Fiber**
21. **Press ONH**
22. Ask the volunteer to look at the **“+”** which should appear slightly on the left of the volunteer for her/his right eye and slightly on the right for her/his left eye

23. Press Auto-Adjust button to enhance the OCT image
24. Make sure the optic nerve head is in the middle of the green circle on the screen below the video screen
25. To adjust click on the centre of the optic nerve on the screen below the video screen
26. Make sure the image is in-between the red lines on the measurement screen
27. Ask the volunteer to blink one last time and then repeat three times “Stare”

28. After the last “Stare” press the capture button on the screen or the black button on top of the joystick

29. **The 3D disk option should appear automatically at this point, if it doesn’t go to Step 1 and select it from the Nerve Fiber menu**

    a. **NB. 3D disk must always done for everyone on their first ever measurement with this machine (not needed for follow up)**

30. Make sure the optic nerve head is in the middle of the green circle on the screen below the video screen

31. To adjust tap with the mouse on the optic nerve on the screen below the video screen

32. Make sure the image is in-between the red lines on the measurement screen

33. Ask the volunteer to blink one last time and then repeat three times “Stare”

34. After the last “Stare” press the capture button on the screen or the black button on top of the joystick

35. Press GCC

36. Advise the volunteer that the target will move to slightly off-centre

37. Press Auto-Adjust button to enhance the OCT image

38. Make sure the optic nerve head is in the middle of the green circle on the screen below the video screen

39. To adjust tap with the mouse on the optic nerve on the screen below the video screen

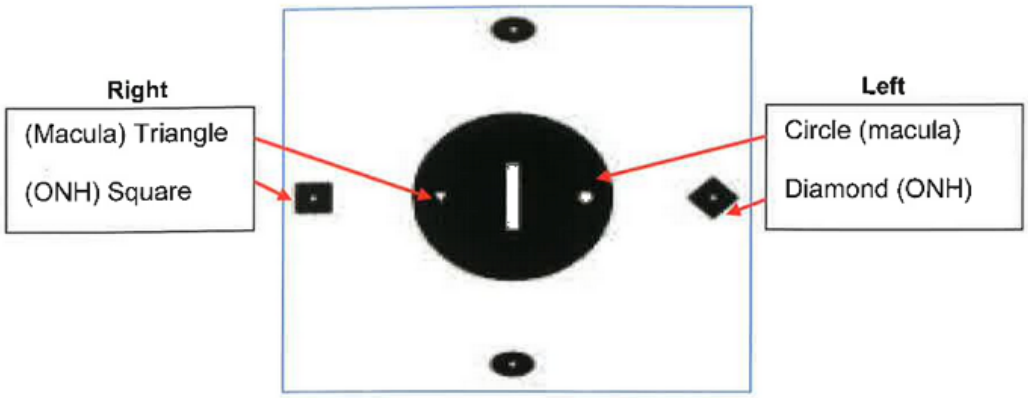
40. Make sure the image is in-between the red lines on the measurement screen

41. Ask the volunteer to blink one last time and then repeat three times “Stare”

42. After the last “Stare” press the capture button on the screen or the black button on top of the joystick

H. Taking Fundus measurements

1. Press the iCam bar on the left
2. Show the volunteer the drawing of the target and explain to them what they will be looking at: When photographing their right eye they should be looking at the small triangle which is in the big circle for the first photo (macula) and at the square for the second photo (ONH). The square will be far on their left (direction their nose) and out of the big circle. When photographing their left eye they should be looking at the small circle which is in the big circle for the first photo. For the second photo they will need to look at the diamond which will be far on their right (direction their nose) and out of the big circle.



3. Press Right/OD
4. Move forward the camera till you see the two bright focus bars. Use the focus knob to align the bars on top of each other
5. Move forward till you see the working distance dots. Move forward and backwards, up and down till they become small and clear and are equally spaced on both sites of the central line
6. Ask the volunteer to look at the triangle
7. Tell the volunteer not to blink
8. Take a photo by pressing capture or the black button on the joystick

9. If you are happy with the photo press save if not press discard and repeat the procedure
10. Ask the volunteer to look at the square
11. Tell the volunteer not to blink
12. Take a photo by pressing capture or the black button on the joystick
13. If you are happy with the photo press save if not press discard and repeat the procedure
14. Press Left/OS
15. Move forward the camera till you see the two bright focus bars. Use the focus knob to align the bars on top of each other
16. Move forward till you see the working distance dots. Move forward and backwards, up and down till they become small and clear and are equally spaced on both sites of the central line
17. Ask the volunteer to look at the small circle
18. Tell the volunteer not to blink
19. Take a photo by pressing capture or the black button on the joystick
20. If you are happy with the photo press save if not press discard and repeat the procedure
21. Ask the volunteer to look at the diamond
22. Tell the volunteer not to blink
23. If you are happy with the photo press save if not press discard and repeat the procedure

I. When finished ask the volunteer if they want to see the results. If the answer is yes guide the volunteer trough the results

1. Go to the review bar and guide the volunteer trough the iCam results first
2. Point out the retina, the optic nerve and the arteries and veins

3. If there is no visible pathology according to your understanding tell to the volunteer one of the two below (dependent on your status):
    a. you are not a trained ophthalmologist but as far as you can see there are no problems. The photos will be checked by a specialist and we will contact the volunteers if we find anything wrong
    b. you are an ophthalmologist and as far as you can see there are no problems
4. If there is time also show the volunteer the results from the OCT: the retina map and ONH only and guide them through these results too
5. If you have observed a problem with the eye that needs attention do one of the options below (dependent on your knowledge):
    a. If you are absolutely confident that you know what the problem is but you are not an ophthalmologist, explain to them that what you found and if needed arrange a letter of referral to be sent to their GP but double-check with the ophthalmologist first.
    b. If you do not know what the problem is explain that you are not a trained ophthalmologist and that you will ask a specialist to look at the results and we will contact them if they need referring or if they need further attention
    c. If you are ophthalmologist explain what is wrong with them and write them referral to the GP immediately if needed
    d. In general we do not refer dry AMD, drusen, epiretinal membrane detachment, small splits. We refer wet AMD and glaucoma, cystoid macular oedema post cataract surgery, haemorrhages, macular holes.

Appendix II: Publications resulting from this thesis

Only article where I am first or stared first author are included in this appendix.

Research

JAMA Ophthalmology | Original Investigation

Genetic and Environmental Factors Associated With the Ganglion Cell Complex in a Healthy Aging British Cohort

Edward Bloch, MBBS; Ekaterina Yonova-Doing, MSc; Eneh Jones-Odeh, MBBS; Katie M. Williams, FRCOphth; Diana Kozareva, BSc; Christopher J. Hammond, MD

IMPORTANCE Measurement of ganglion cell complex (GCC) thickness may be more sensitive than current methods for glaucoma diagnosis and research. However, little is known about the factors influencing GCC thickness in the general population.

OBJECTIVES To investigate the heritability of and factors associated with GCC thickness in a healthy aging population.

DESIGN, SETTING, AND PARTICIPANTS A cross-sectional twin study was conducted from August 27, 2014, to March 31, 2016, among 1657 participants of white British ancestry from the TwinsUK study cohort without ocular pathologic conditions. Heritability analyses were conducted in 1432 twins (426 monozygous and 290 dizygous pairs). Association analyses were performed using univariable and multivariable stepwise linear regression models, taking family structure into account. Heritability analyses were conducted using maximum likelihood structural equation twin modeling.

MAIN OUTCOMES AND MEASURES Parameters measured included GCC thickness, autorefraction, intraocular pressure, blood pressure, body mass index, and cholesterol, creatinine, glucose, insulin, triglycerides, and urea levels. Estimated glomerular filtration rate was calculated using the Modification of Diet in Renal Disease formula.

RESULTS Among the 1657 participants (mean [SD] age, 56.0 [15.3] years; 89.5% women and 10.5% men), the mean [SD] inner GCC thickness was 96.0 [7.6] µm (95% CI, 95.1-96.2). In multivariable modeling, the mean inner GCC thickness was associated with advancing age (β, −0.14; $P < .001$), increased body mass index (β, −0.15; $P = .001$), spherical equivalent (β, 0.70; $P < .001$), and higher estimated glomerular filtration rate (β, 0.03; $P = .02$). A 1-U increase in age or body mass index was associated with a 0.14-µm and 0.15-µm decrease in GCC thickness, respectively ($P < .001$), while a 1-U increase in spherical equivalent or estimated glomerular filtration rate was associated with a 0.70-µm ($P < .001$) and 0.03-µm ($P = .02$) increase in GCC thickness, respectively. Ganglion cell complex thickness was not associated with sex, intraocular pressure, or diabetes. Age-adjusted GCC thickness was highly heritable, with additive genetic effects explaining 81% (95% CI, 78%-84%) of phenotypic variance and individual environmental factors explaining the remaining 19% (95% CI, 16%-22%).

CONCLUSIONS AND RELEVANCE Ganglion cell complex thickness appears to be highly heritable and further genetic analysis may help identify new biological pathways for glaucoma. The results suggest it may be important to account for age, body mass index, refractive error, and sex when using GCC thickness as a diagnostic tool. Replication of their results is required, as is further research to explain the association between renal function and GCC thickness.

JAMA Ophthalmol. 2017;135(1):31-38. doi:10.1001/jamaophthalmol.2016.4486
Published online November 17, 2016.

Author Affiliations: Department of Twin Research and Genetic Epidemiology, King's College London, London, England (Bloch, Yonova-Doing, Jones-Odeh, Williams, Kozareva, Hammond); Department of Ophthalmology, King's College London, London, England (Williams, Hammond).

Corresponding Author: Christopher J. Hammond, MD, Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, Third Floor Block D South Wing, St Thomas' Hospital, London SE1 7EH, England (chris.hammond@kcl.ac.uk).

Glaucoma is the second leading cause of blindness worldwide, affecting up to 3% of the global population older than 40 years, and is forecasted to affect 79.6 million people worldwide by 2020.[1,2] Glaucoma comprises a group of optic neuropathies characterized by progressive, irreversible visual field loss. Impaired retrograde neurotrophic transport causes dysfunction of retinal ganglion cells and, ultimately, apoptosis with axonal atrophy.[3-5]

As loss of retinal ganglion cells is not visible on ophthalmoscopy, glaucoma is traditionally detected using optic nerve examination and visual field tests. These tests have low sensitivity, are prone to interobserver variation, and detect changes that occur relatively late in the disease process, leading to a diagnostic delay of up to 10 years.[6-9] The advent of spectral domain-optical coherence tomography (SD-OCT) has afforded the opportunity to acquire noninvasive in vivo high-resolution segmentation of the inner retinal layers.[10-12] Since the macula contains up to 50% of retinal ganglion cells, it is very sensitive to early glaucomatous damage.[13] Spectral domain-optical coherence tomography measurements of the macular ganglion cell complex (GCC) thickness, comprising the ganglion cell layer, the inner plexiform layer, and the retinal nerve fiber layer (RNFL), correlate closely with histologic data.[14] Various studies have investigated the diagnostic ability and validity of SD-OCT parameters, concluding that both GCC and peripapillary RNFL thickness have high sensitivity, specificity, and positive predictive value for disease.[15-27]

It is well established that older age and higher intraocular pressure (IOP) are risk factors for primary open-angle glaucoma (POAG).[28] It also has been suggested that factors such as vascular dysregulation[29-32] and diabetes[33] may play an important role in the development of POAG. In addition, family history is known to contribute to the risk of developing POAG[28] and genetic variants have been identified that play a role in POAG and its endophenotypes.[34,35]

The extent to which any of these factors may influence GCC thickness in a healthy population is unclear. We undertook a cross-sectional observational study to explore the systemic and ophthalmic factors influencing GCC thickness in a healthy population of European ancestry. In particular, we focused on parameters that might be associated with microvascular damage, such as blood pressure, diabetes, and renal function. To investigate to what extent genetic factors influence GCC thickness, we used a twin model design to perform a heritability study.

Key Points

Question What are the genetic and environmental factors associated with ganglion cell complex thickness in a healthy aging British cohort?

Findings In a cross-sectional twin cohort study of 1657 participants, ganglion cell complex thickness was associated with age, body mass index, and refractive error, as well as estimated glomerular filtration rate. Ganglion cell complex thickness was highly heritable, with additive genetic effects explaining 81% of phenotypic variance.

Meaning Ganglion cell complex thickness appears to be a highly heritable trait and adjustment for age, body mass index, and refractive error is important when using it as a diagnostic parameter.

Methods

Participants

All monozygous and dizygous twin pairs were volunteers in the TwinsUK study, an unselected population-based twin cohort representative of the broader population in terms of disease-associated and lifestyle characteristics.[36] We assessed 1753 individuals for eligibility and included those of European ancestry and who were 18 years or older and excluded 39 not of European ancestry, 33 with treated glaucoma, 4 with ocular hypertension, 7 with other pathologic conditions affecting the retina, 12 with poor-quality SD-OCT images, and 1 with incorrect data. Zygosity was determined by a standardized questionnaire and confirmed using genome-wide genotyping of single-nucleotide polymorphisms or short tandem repeats in some monozygous twin pairs for whom genome-wide data were not available. The study was approved by the Guy's and St. Thomas' ethics committee and all the participants provided written informed consent in accordance with the Declaration of Helsinki.[37]

Clinical Examination

Data were collected between August 27, 2014, and March 31, 2016. Each participant completed a questionnaire to elicit any previous ophthalmic history. Nonophthalmic parameters were assessed, including height, weight, blood pressure, and standard blood markers relevant to microvascular disease (ie, levels of cholesterol, creatinine, glucose, insulin, triglycerides, and urea). Estimated glomerular filtration rate (eGFR) was calculated using the Modification of Diet in Renal Disease formula.[38]

Ophthalmic examination on both eyes included autorefraction, pachymetry, and noncontact measurement of IOP (Visionix120; The Luneau Technology Group). Spherical equivalent (SphE) was calculated using the standard formula of sphere + (cylinder/2). Fundus photography and SD-OCT, including mapping of GCC thickness, were performed (iVue SD-OCT; Optovue).[18,19,39-41] Scans were reviewed by an ophthalmologist (E.B.) and poor-quality scans (n = 59, confirmed by K.M.W.) were excluded from the final analyses. As there was a high correlation between left and right eyes for all measurements (Pearson correlation coefficient, >0.7), we used the mean of the 2 eyes when both eyes were available.

Statistical Analysis

Differences between monozygous and dizygous twins, or between any other group (eg, participants with hypertension or diabetes vs controls), were compared using 2-sample, 2-tailed *t* tests or *z* tests, assuming equal variance. Associations were assessed between GCC thickness and the following 14 factors: age, sex, body mass index (BMI; calculated as weight in kilograms divided by height in meters squared), diabetes status, blood pressure (systolic and diastolic), 2 ophthalmic parameters (SphE and IOP), and 6 blood markers (levels of cholesterol, creatinine, glucose, insulin, triglycerides, and urea).

Univariable linear regression analyses were performed, followed by a multivariable linear regression model for factors with an association significant at $P < .05$ in the univariable model. Independent variables were identified using stepwise backward procedure with a threshold for removal set at $P < .05$. Variables that, in the multivariable model, would survive Bonferroni correction for multiple testing ($P \leq .004$; 0.05 divided by 14, where 14 is the number of independent variables tested), were considered associated with GCC thickness. Mean arterial pressure and eGFR are not independent variables, as they are calculated from these variables, so they do not affect the Bonferroni correction threshold. In all regression models, family structure was taken into account.

All analyses, except for the heritability, were carried out using STATA, version 14, statistical package (StataCorp). Heritability analyses were performed using maximum likelihood structural equation twin modeling implemented in the OpenMx package in R (http://openmx.psyc.virginia.edu). We compared the phenotypic variance between monozygous and dizygous twins to estimate the extent to which GCC thickness was a result of additive or dominant genetic effects and common or unique environmental factors. The goodness of fit of the full and reduced additive, common, and unique (ACE) or additive, dominant, and unique (ADE) models was compared with the observed data and the best-fitting model was selected. Before the heritability analysis, GCC thickness was adjusted for age by linear regression of GCC and age and applying the models to the residuals.

Results

Data on GCC thickness were acquired from 3235 eyes of 1657 predominantly female (89.5%) participants, whose demographic and clinical characteristics are presented in **Table 1**. We excluded 39 individuals of non-European ancestry, 44 with ocular pathologic conditions, and 13 for other reasons. The individuals who were excluded for ocular pathologic conditions or other reasons were a mean of 14.3 years older than the remaining participants, who were a mean (SD) age of 56.0 (15.3) years (range, 18-90 years). Monozygous twins were on average younger than the dizygous twins (54 vs 59 years) and had thicker mean GCC (96.4 vs 95.3 µm; $P = .01$). Monozygous and dizygous twins also had different mean (SD) values for the following parameters: BMI (25.6 [0.16] vs 26.5 [0.21]), SphE (-0.21 [0.08] vs 0.15 [0.10] diopters), systolic blood pressure (126.3 [0.58] vs 128.9 [0.66] mm Hg), and serum cholesterol level (206.95 [1.54] vs 212.36 [1.54] mg/dL [to convert to millimoles per liter, multiply by 0.0259]) ($P \leq .01$ for all). After linear age adjustment, there were no differences between monozygous and dizygous twins for these variables.

The mean (SD) inner GCC thickness of the study population was 96.0 (7.6) µm (95% CI, 95.1-96.2), with mean (SD) superior thickness of 95.4 (7.7) µm (95% CI, 95.1-95.7) and inferior thickness of 96.6 (8.0) µm (95% CI, 96.2-96.8). The superior and inferior thicknesses were highly correlated with the mean inner GCC thickness (Pearson correlation coefficient, 0.97 for both) and yielded similar results in all analyses.

Table 1. Demographic and Clinical Characteristics of the Study Population

Characteristic	Value
Age, mean (SD), y	56.0 (15.3)
Height, mean (SD), m	
Women	1.6 (0.1)
Men	1.7 (0.1)
Weight, mean (SD), kg	
Women	68.6 (13.2)
Men	82.6 (16.4)
Body mass index, mean (SD)[a]	26.2 (5.1)
Women	26.1 (4.9)
Men	27.3 (4.5)
Blood pressure, mean (SD), mm Hg	
Systolic	127.3 (17.4)
Diastolic	75.8 (10.6)
Female sex, No. (%)	1483 (89.5)
Diabetes, No. (%)	190 (11.5)
Hypertension, No. (%)	290 (17.5)
Receiving antihypertensive treatment, No. (%)	213 (12.9)
IOP, mean (SD), mm Hg	13.3 (2.6)
SphE, mean (SD), diopters	−0.1 (2.4)
CCT, mean (SD), µm	536.2 (35.6)
Total cholesterol, mean (SD), mg/dL	208.5 (42.5)
Creatinine, mean (SD), mg/dL	0.83 (0.15)
Glucose, mean (SD), mg/dL	86.5 (10.8)
Insulin, mean (SD), µIU/mL	49.1 (32.2)
Total triglycerides, mean (SD), mg/dL	88.5 (53.1)
Urea, mean (SD), mg/dL	32.43 (6.6)

Abbreviations: CCT, central corneal thickness; IOP, intraocular pressure; SphE, spherical equivalent.

SI conversion factors: To convert cholesterol to millimoles per liter, multiply by 0.0259; creatinine to micromoles per liter, multiply by 88.4; glucose to millimoles per liter, multiply by 0.0555; triglycerides to millimoles per liter, multiply by 0.0113; and urea to micromoles per liter, multiply by 59.485.

[a] Calculated as weight in kilograms divided by height in meters squared.

Table 2 summarizes the results from the univariable and multivariable stepwise regression analyses. On univariable analysis, the mean GCC was thinner with advancing age (β [SE], -0.12 [0.02]; $P < .001$), increased BMI (β [SE], -0.19 [0.04]; $P < .001$), and higher systolic (β [SE], -0.06 [0.01]; $P < .001$) and diastolic blood pressures (β [SE], -0.06 [0.02]; $P = .003$). The association with systolic and diastolic blood pressure survived adjustment for hypertension status (β [SE], -0.05 [0.01]; $P < .001$; and β [SE], -0.04 [0.02]; $P = .02$, respectively). Of the ocular parameters, only SphE was associated with GCC thickness on univariable analysis (β [SE], 0.45 [0.09]; $P < .001$), with a thinner GCC associated with a more myopic refraction. Regarding blood parameters, on univariable analysis GCC thickness was negatively associated with an increase in levels of creatinine (β [SE], -2.69 [1.28]; $P = .01$), glucose (β [SE], -0.27 [0.003]; $P = .01$), insulin (β [SE], -0.02 [0.01]; $P = .01$), triglyceride (β [SE], -0.01 [0.004]; $P = .01$), and urea (β [SE], -3.42 [0.90]; $P < .001$) levels (Table 2). In multivariable modeling, the mean inner GCC thickness was independently associated with advancing age (β [SE], -0.14 [0.02]; $P < .001$), SphE (β [SE],

Table 2. Univariable and Multivariable Linear Regression Analysis

Characteristic	β (SE) Univariable	*P* Value	Multivariable	*P* Value
Age	−0.12 (0.02)	<.001	−0.14 (0.02)	<.001
Sex	0.31 (0.77)	.69		
Body mass index	−0.19 (0.04)	<.001	−0.15 (0.04)	.001
Blood pressure				
Systolic	−0.06 (0.01)	<.001		
Diastolic	−0.06 (0.02)	.003		
Diabetes	−0.91 (0.47)	.05		
Spherical equivalent	0.45 (0.09)	<.001	0.70 (0.09)	<.001
Intraocular pressure	0.05 (0.08)	.51		
Total cholesterol	−0.01 (0.01)	.27		
Creatinine	−2.69 (1.28)	.01		
Glucose	−0.27 (0.03)	.01		
Insulin	−0.02 (0.01)	.01		
Total triglycerides	−0.01 (0.004)	.01		
Urea	−3.42 (0.90)	<.001		

0.70 [0.09]; $P < .001$), and increased BMI (β [SE], -0.15 [0.04]; $P = .001$) only; these associations would survive correction for multiple testing (Bonferroni threshold, $P \leq .004$).

Given the possible connection between GCC thickness and both blood pressure[42-44] and kidney function,[45] we explored associated parameters even though they did not survive the multivariable analysis. We first examined if mean arterial pressure, which is considered to be a better marker of organ perfusion,[46] was associated with GCC thickness. The mean arterial pressure in our study was a mean (SD) of 92.9 (12.1) mm Hg. Similar to systolic and diastolic blood pressure, mean arterial pressure was associated with GCC thickness, irrespective of hypertension status (β, -0.06; $P < .001$), but not when adjusted for age (β, -0.03; $P = .13$). Next, we looked at whether GCC thickness was associated with the eGFR. In our study, the mean (SD) eGFR was 82.5 (19.6) mL/min/1.73m^2. Higher eGFR was associated with a thicker GCC, independent of age, hypertension, and diabetes status (β, 0.03; $P = .02$).

Finally, we calculated the heritability of GCC thickness in 1432 twins (426 monozygous and 290 dizygous pairs), as 225 individuals had their twin missing. We found the best-fitting model to be the additive genetic and unique environment (AE) model, in which 81% (95% CI, 78%-84%) of variance in age-adjusted GCC thickness was explained by additive genetic factors and the remaining 19% (95% CI, 16%-22%) by unique environmental factors.

Discussion

Many cross-sectional and longitudinal studies have investigated the effects of age, sex, race/ethnicity, IOP, and axial length on RNFL thickness.[47-54] As GCC incorporates RNFL, the 2 variables are correlated (Pearson correlation coefficient, 0.76 in this sample) and likely share risk factors. However, we set out to examine the associations in a large population-based epidemiologic study of GCC thickness in healthy European individuals and found it to be strongly and independently associated with age, BMI, SphE, and eGFR. Except for eGFR, all the other parameters survived correction for multiple testing.

With respect to age, our results are similar to those previously found for RNFL.[55-57] We found, cross-sectionally, a thinner GCC with older age (a mean decrease of 0.14 µm per year). The effect we found is comparable with that found by Zhang et al[54] in 92 controls (0.17 µm per year) in the Advanced Imaging for Glaucoma Study. In contrast, the association between RNFL and BMI is more variable. Morbid obesity had a significant influence on RNFL, retinal ganglion cells, and choroidal thickness in 1 study,[58] while others suggest that BMI plays a role in RNFL thickness only in men[55] or not at all.[59] The UK Biobank study found a significant negative association between macular thickness and both BMI and male sex,[56] which is in keeping with our study findings of a reduction in GCC thickness of 0.15 µm per 1-U increase in BMI. Owing to the small number of male participants in this study, we were underpowered to detect sex effects; however, the direction of effect was the same in the 2 groups (β, -0.19 and $P < .001$ for women; β, -0.28 and $P = .20$ for men).

Studies of RNFL thickness from the European Prospective Investigation Into Cancer and Nutrition (EPIC)-Norfolk and UK Biobank cohorts found RNFL and macular thickness to be thinner in men than in women.[55,56] Wang et al[60] found no sex differences in the mean thickness of GCC and ganglion cells and the inner plexiform layer in a Chinese cohort. However, while Mwanza et al[61] did not find a difference between the sexes for mean thickness of ganglion cells and the inner plexiform layer ($P = .36$), they did find that male sex was associated with GCC thickness in multivariate analysis with RNFL thickness, age, and axial length (β, -1.62; $P = .005$). We did not observe a difference in mean (SD) GCC thickness between the 174 men (95.7 [7.98] µm) and 1483 women (96.1 [7.45] µm; $P = .69$) in our study.

Of the ophthalmic parameters, GCC thickness was associated only with SphE, with more myopic eyes having a thinner GCC and a mean increase in thickness of 0.70 µm per 1-diopter increase in SphE. Previous studies have proposed that longer axial length is associated with thinner RNFL.[62-64] However, Khawaja et al,[55] using scanning laser polarimetry in the EPIC-Norfolk study, found that longer axial length appeared to be associated with a thicker RNFL. There is some controversy as to the possible magnification effects of the eye with increasing ametropia, which is not routinely corrected for by SD-OCT software.[55,65,66] Some studies using SD-OCT have found that this association reversed once ocular magnification was mathematically corrected.[65,67] Patel et al[56] demonstrated that, on results of SD-OCT, increasing myopia was associated with increased central macular thickness but a decrease in the other macular subfield thicknesses, resulting in an overall positive association with refractive error. Optical coherence tomography also has been shown to be more repeatable and sensitive than glaucoma diagnoses by scanning laser polarimetry at detecting glaucoma in cases of high myopia.[40,68-72] Irrespective of whether the effect is physiological or owing to optics, we propose that future studies of GCC

should consider adjusting for SphE or axial length, given their strong association with GCC thickness.

Although IOP is a known risk factor for POAG, we were not able to identify an association between IOP and GCC thickness in our study. However, this is in agreement with the findings from RNFL association studies, such as the EPIC-Norfolk, UK Biobank, and Advanced Imaging for Glaucoma Study groups.[54-56] It is possible that our exclusion of individuals taking IOP-lowering medication diminished the association with GCC thickness, although we excluded only 37 individuals on this criterion. Furthermore, there is increasing evidence for the role of the association between IOP and intracranial pressure, known as the translaminar pressure difference, in the pathophysiologic cause of glaucoma.[73,74] It may be that the IOP does not damage retinal ganglion cell axons at the lamina cribrosa until it reaches a given threshold, which would, in part, further explain the pathophysiologic cause of normal-tension glaucoma.

Vascular dysregulation has been proposed to play an important role in POAG[29-32] and several recent studies have suggested that systemic hypertension is associated with RNFL and GCC thinning, which may represent an important consideration when using retinal SD-OCT measurements as a diagnostic tool.[42-44] Therefore, in this study, we explored the association between GCC thickness and blood pressure (systolic, diastolic, and mean arterial pressure). We did find GCC to be thinner in individuals with hypertension (94.3 µm) vs those with normal blood pressure (96.3 µm; $P < .001$), and an increase in any of the blood pressure parameters was associated with GCC thickness in univariable analyses. However, this association was small (β, -0.06 µm) and not independent from the effect of age.

Similarly, some studies have shown diabetes to have an early neurodegenerative effect on the retina, causing neuronal dysfunction, leading to thinning of both the RNFL and GCC, with associated functional visual deficits.[75-82] Chhablani et al[79] showed that inner retinal thinning was present in patients with type 2 diabetes, even before signs of retinopathy were visible, while Salvi et al[80] showed a significant association of GCC thinning in patients with diabetic polyneuropathy but not in those with retinopathy. We did not find an association between a diagnosis of diabetes and GCC thickness, nor with fasting serum glucose and insulin levels. The participants in the TwinsUK study cohort are generally healthy, so we may be underpowered to detect effects specific to hypertension or diabetes, although our study did include 290 and 190 participants with these respective diagnoses.

Finally, we found eGFR to be associated with mean GCC thickness. The association survived correction for the effects of age, diabetes, and hypertension, accounting for a 0.03-µm change in GCC thickness per 1-U change in eGFR. This finding suggests that there may be an association between renal function and retinal neuropathy. Srivastav et al[83] found an association between increased serum urea and creatinine levels and decreased RNFL thickness in patients with diabetes, while Shim et al[84] recently demonstrated that low eGFR levels are independently associated with POAG. However, there is a paucity of data regarding this potential association in the absence of diabetes. One study, by Demir et al,[45] evaluated the RNFL thickness in patients with chronic renal failure without diabetes, concluding that it was significantly thinner than normal. Our data are consistent with this notion, but further investigation is required to elucidate the underlying association. It is accepted that the kidneys and retina share common developmental pathways and structural similarities, such as the vascular configuration and type IV collagen basement-membrane composition.[85] Therefore, if renal function correlates with retinal neuronal anatomy in a healthy population, it may become a factor to consider in the interpretation of GCC thickness as a diagnostic parameter.

This is the first study, to our knowledge, that explores the heritability of GCC thickness. We found GCC thickness to be highly heritable, with additive genetic effects accounting for 81% of variance, while unique environmental effects explained the remaining 19% of variance. This finding is comparable with the heritability estimates for central retinal thickness (90%) and RNFL thickness (between 48% and 82%).[86-88] It would be interesting to explore whether any of the known genetic risk factors for POAG and its other endophenotypes also influence GCC thickness. Alleles in *SIX1-SIX6* (GenBank: 6495 and 4990, respectively) have been associated with POAG in genome-wide association studies, and with a decrease in global and sectoral RNFL thickness in individuals of European and Asian descent.[89-91]

Limitations

This study contains some limitations. First, the TwinsUK study cohort is predominantly females of European ancestry, limiting our ability to draw conclusions as to the effect of sex or race/ethnicity on GCC thickness. As this is a cross-sectional study, we cannot draw any conclusions on causation and, as an observational study, we are potentially susceptible to residual confounding and missing data. As we relied on self-reported ocular and systemic disease, as well as drug history, we may be liable to misspecification of variables. Our novel findings require replication.

Conclusions

Ganglion cell complex thickness is a highly heritable trait. We identified an association between SD-OCT-derived GCC thickness measurements and age, refractive error, BMI, and eGFR in a healthy, aging, predominantly female cohort of European descent. The associations between SphE, BMI, and age are strong and therefore adjustment is likely to be relevant if GCC thickness were to be used as a clinical parameter. Future studies of GCC and its association with glaucoma or neurologic disease should take these associations into account. We did not find GCC thickness to be independently associated with sex, diabetes, blood pressure, or IOP. Further studies are required to establish the precise association between renal function and inner retinal structures and to investigate the genetic factors underlying GCC thickness.

ARTICLE INFORMATION

Accepted for Publication: September 28, 2016.

Published Online: November 17, 2016. doi:10.1001/jamaophthalmol.2016.4486

Author Contributions: Dr Bloch and Ms Yonova-Doing had full access to all the data in the study and take full responsibility for the integrity of the data and the accuracy of the data analysis. Dr Bloch and Ms Yonova-Doing contributed equally to this work.
Study concept and design: Bloch, Yonova-Doing, Williams, Hammond.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Bloch, Yonova-Doing, Hammond.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Bloch, Yonova-Doing, Jones-Odeh, Williams, Hammond.
Obtained funding: Hammond.
Administrative, technical, or material support: Bloch, Williams, Kozareva.
Study supervision: Hammond.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This study received funding from the International Glaucoma Association. The Wellcome Trust and the European Community's Seventh Framework Programme (FP7/2007-2013) supported the TwinsUK study. The study also received support from the National Institute for Health Research–funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St. Thomas' National Health Service Foundation Trust in partnership with King's College London. Ms Yonova-Doing received funding from the Biotechnology and Biological Sciences Research Council and the London Interdisciplinary Doctoral Program. Ms Williams received funding from the Medical Research Council Clinical Research Training Fellowship.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank all the twin volunteers for their participation.

REFERENCES

1. Varma R, Lee PP, Goldberg I, Kotak S. An assessment of the health and economic burdens of glaucoma. *Am J Ophthalmol*. 2011;152(4):515-522.

2. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90(3):262-267.

3. Quigley HA. Neuronal death in glaucoma. *Prog Retin Eye Res*. 1999;18(1):39-57.

4. McKinnon SJ. Glaucoma, apoptosis, and neuroprotection. *Curr Opin Ophthalmol*. 1997;8(2):28-37.

5. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311(18):1901-1911.

6. Harwerth RS, Carter-Dawson L, Shen F, Smith EL III, Crawford ML. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Invest Ophthalmol Vis Sci*. 1999;40(10):2242-2250.

7. Cordeiro MF, Guo L, Luong V, et al. Real-time imaging of single nerve cell apoptosis in retinal neurodegeneration. *Proc Natl Acad Sci U S A*. 2004;101(36):13352-13356.

8. Balendra SI, Normando EM, Bloom PA, Cordeiro MF. Advances in retinal ganglion cell imaging. *Eye (Lond)*. 2015;29(10):1260-1269.

9. Tatham AJ, Weinreb RN, Zangwill LM, Liebmann JM, Girkin CA, Medeiros FA. The relationship between cup-to-disc ratio and estimated number of retinal ganglion cells. *Invest Ophthalmol Vis Sci*. 2013;54(5):3205-3214.

10. Wang M, Hood DC, Cho JS, et al. Measurement of local retinal ganglion cell layer thickness in patients with glaucoma using frequency-domain optical coherence tomography [published correction appears in *Arch Ophthalmol*. 2010;128(9):1150]. *Arch Ophthalmol*. 2009;127(7):875-881.

11. Ishikawa H, Stein DM, Wollstein G, Beaton S, Fujimoto JG, Schuman JS. Macular segmentation with optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2005;46(6):2012-2017.

12. Chen TC, Cense B, Pierce MC, et al. Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging. *Arch Ophthalmol*. 2005;123(12):1715-1720.

13. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol*. 1990;300(1):5-25.

14. Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1-21.

15. Yang Z, Tatham AJ, Weinreb RN, Medeiros FA, Liu T, Zangwill LM. Diagnostic ability of macular ganglion cell inner plexiform layer measurements in glaucoma using swept source and spectral domain optical coherence tomography. *PLoS One*. 2015;10(5):e0125957.

16. Tan O, Chopra V, Lu AT, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology*. 2009;116(12):2305-14.e1, 2.

17. Kim NR, Lee ES, Seong GJ, Kim JH, An HG, Kim CY. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. *Invest Ophthalmol Vis Sci*. 2010;51(9):4646-4651.

18. Schulze A, Lamparter J, Pfeiffer N, Berisha F, Schmidtmann I, Hoffmann EM. Diagnostic ability of retinal ganglion cell complex, retinal nerve fiber layer, and optic nerve head measurements by Fourier-domain optical coherence tomography. *Graefes Arch Clin Exp Ophthalmol*. 2011;249(7):1039-1045.

19. Garas A, Vargha P, Holló G. Diagnostic accuracy of nerve fibre layer, macular thickness and optic disc measurements made with the RTVue-100 optical coherence tomograph to detect glaucoma. *Eye (Lond)*. 2011;25(1):57-65.

20. Medeiros FA, Zangwill LM, Bowd C, Vessani RM, Susanna R Jr, Weinreb RN. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am J Ophthalmol*. 2005;139(1):44-55.

21. Tan O, Li G, Lu AT, Varma R, Huang D; Advanced Imaging for Glaucoma Study Group. Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis. *Ophthalmology*. 2008;115(6):949-956.

22. Jeoung JW, Choi YJ, Park KH, Kim DM. Macular ganglion cell imaging study: glaucoma diagnostic accuracy of spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2013;54(7):4422-4429.

23. Mwanza JC, Durbin MK, Budenz DL, et al. Glaucoma diagnostic accuracy of ganglion cell-inner plexiform layer thickness: comparison with nerve fiber layer and optic nerve head. *Ophthalmology*. 2012;119(6):1151-1158.

24. Oddone F, Lucenteforte E, Michelessi M, et al. Macular versus retinal nerve fiber layer parameters for diagnosing manifest glaucoma: a systematic review of diagnostic accuracy studies. *Ophthalmology*. 2016;123(5):939-949.

25. Loewen NA, Zhang X, Tan O, et al; Advanced Imaging for Glaucoma Study Group. Combining measurements from three anatomical areas for glaucoma diagnosis using Fourier-domain optical coherence tomography. *Br J Ophthalmol*. 2015;99(9):1224-1229.

26. Nouri-Mahdavi K, Nowroozizadeh S, Nassiri N, et al. Macular ganglion cell/inner plexiform layer measurements by spectral domain optical coherence tomography for detection of early glaucoma and comparison to retinal nerve fiber layer measurements. *Am J Ophthalmol*. 2013;156(6):1297-1307.e2.

27. Meira-Freitas D, Lisboa R, Tatham A, et al. Predicting progression in glaucoma suspects with longitudinal estimates of retinal ganglion cell counts. *Invest Ophthalmol Vis Sci*. 2013;54(6):4174-4183.

28. Coleman AL, Miglior S. Risk factors for glaucoma onset and progression. *Surv Ophthalmol*. 2008;53(suppl 1):S3-S10.

29. Osborne NN, Melena J, Chidlow G, Wood JP. A hypothesis to explain ganglion cell death caused by vascular insults at the optic nerve head: possible implication for the treatment of glaucoma. *Br J Ophthalmol*. 2001;85(10):1252-1259.

30. Moore D, Harris A, Wudunn D, Kheradiya N, Siesky B. Dysfunctional regulation of ocular blood flow: a risk factor for glaucoma? *Clin Ophthalmol*. 2008;2(4):849-861.

31. Hayreh SS. The role of age and cardiovascular disease in glaucomatous optic neuropathy. *Surv Ophthalmol*. 1999;43(suppl 1):S27-S42.

32. Lawrenson JG. Histopathology and pathogenesis of glaucomatous optic neuropathy. In: Edgar DF, Rudnicka AR, eds. *Glaucoma Identification and Co-management*. Philadelphia, PA: Elsevier Limited; 2007:27-35.

33. Zhou M, Wang W, Huang W, Zhang X. Diabetes mellitus as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. *PLoS One*. 2014;9(8):e102972.

34. Charlesworth J, Kramer PL, Dyer T, et al. The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness. *Invest Ophthalmol Vis Sci*. 2010;51(7):3509-3514.

35. Abu-Amero K, Kondkar AA, Chalam KV. An updated review on the genetics of primary open angle glaucoma. *Int J Mol Sci*. 2015;16(12):28886-28911.

36. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol*. 2013;42(1):76-85.

37. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053

38. Levey AS, Coresh J, Greene T, et al; Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145(4):247-254.

39. Mori S, Hangai M, Sakamoto A, Yoshimura N. Spectral-domain optical coherence tomography measurement of macular volume for diagnosing glaucoma. *J Glaucoma*. 2010;19(8):528-534.

40. Bertuzzi F, Benatti E, Esempio G, Rulli E, Miglior S. Evaluation of retinal nerve fiber layer thickness measurements for glaucoma detection: GDx ECC versus spectral-domain OCT. *J Glaucoma*. 2014;23(4):232-239.

41. Takagi ST, Kita Y, Takeyama A, Tomita G. Macular retinal ganglion cell complex thickness and its relationship to the optic nerve head topography in glaucomatous eyes with hemifield defects. *J Ophthalmol*. 2011;2011:914250.

42. Gangwani RA, Lee JW, Mo HY, et al. The correlation of retinal nerve fiber layer thickness with blood pressure in a Chinese hypertensive population. *Medicine (Baltimore)*. 2015;94(23):e947.

43. Akay F, Gündoğan FC, Yolcu U, Toyran S, Tunç E, Uzun S. Retinal structural changes in systemic arterial hypertension: an OCT study. *Eur J Ophthalmol*. 2016;26(5):436-441.

44. Sahin OZ, Sahin SB, Ayaz T, et al. The impact of hypertension on retinal nerve fiber layer thickness and its association with carotid intima media thickness. *Blood Press*. 2015;24(3):178-184.

45. Demir MN, Eksioglu U, Altay M, et al. Retinal nerve fiber layer thickness in chronic renal failure without diabetes mellitus. *Eur J Ophthalmol*. 2009;19(6):1034-1038.

46. Brzezinski WA. Blood pressure. In: Walker HK, Hall WD, Hurst JW, eds. *Clinical Methods: The History, Physical and Laboratory Examination*. 3rd ed. Boston, MA: Butterworths; 1990:95-97.

47. Zhao L, Wang Y, Chen CX, Xu L, Jonas JB. Retinal nerve fibre layer thickness measured by Spectralis spectral-domain optical coherence tomography: the Beijing Eye Study. *Acta Ophthalmol*. 2014;92(1):e35-e41.

48. Patel NB, Lim M, Gajjar A, Evans KB, Harwerth RS. Age-associated changes in the retinal nerve fiber layer and optic nerve head. *Invest Ophthalmol Vis Sci*. 2014;55(8):5134-5143.

49. Girkin CA, McGwin G Jr, Sinai MJ, et al. Variation in optic nerve and macular structure with age and race with spectral-domain optical coherence tomography. *Ophthalmology*. 2011;118(12):2403-2408.

50. Budenz DL, Anderson DR, Varma R, et al. Determinants of normal retinal nerve fiber layer thickness measured by Stratus OCT. *Ophthalmology*. 2007;114(6):1046-1052.

51. Leung CK, Yu M, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a prospective analysis of age-related loss. *Ophthalmology*. 2012;119(4):731-737.

52. Leung CK, Ye C, Weinreb RN, Yu M, Lai G, Lam DS. Impact of age-related change of retinal nerve fiber layer and macular thicknesses on evaluation of glaucoma progression. *Ophthalmology*. 2013;120(12):2485-2492.

53. Ueda K, Kanamori A, Akashi A, Tomioka M, Kawaka Y, Nakamura M. Effects of axial length and age on circumpapillary retinal nerve fiber layer and inner macular parameters measured by 3 types of SD-OCT instruments. *J Glaucoma*. 2016;25(4):383-389.

54. Zhang X, Francis BA, Dastiridou A, et al; Advanced Imaging for Glaucoma Study Group. Longitudinal and cross-sectional analyses of age effects on retinal nerve fiber layer and ganglion cell complex thickness by Fourier-Domain OCT. *Transl Vis Sci Technol*. 2016;5(2):1.

55. Khawaja AP, Chan MP, Garway-Heath DF, et al. Associations with retinal nerve fiber layer measures in the EPIC-Norfolk Eye Study. *Invest Ophthalmol Vis Sci*. 2013;54(7):5028-5034.

56. Patel PJ, Foster PJ, Grossi CM, et al; UK Biobank Eyes and Vision Consortium. Spectral-domain optical coherence tomography imaging in 67 321 adults: associations with macular thickness in the UK Biobank study. *Ophthalmology*. 2016;123(4):829-840.

57. Knight OJ, Girkin CA, Budenz DL, Durbin MK, Feuer WJ; Cirrus OCT Normative Database Study Group. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT. *Arch Ophthalmol*. 2012;130(3):312-318.

58. Dogan B, Kazim Erol M, Dogan U, et al. The retinal nerve fiber layer, choroidal thickness, and central macular thickness in morbid obesity: an evaluation using spectral-domain optical coherence tomography. *Eur Rev Med Pharmacol Sci*. 2016;20(5):886-891.

59. Zhu BD, Li SM, Li H, et al; Anyang Childhood Eye Study Group. Retinal nerve fiber layer thickness in a population of 12-year-old children in central China measured by iVue-100 spectral-domain optical coherence tomography: the Anyang Childhood Eye Study. *Invest Ophthalmol Vis Sci*. 2013;54(13):8104-8111.

60. Wang J, Gao X, Huang W, et al. Swept-source optical coherence tomography imaging of macular retinal and choroidal structures in healthy eyes. *BMC Ophthalmol*. 2015;15(15):122.

61. Mwanza JC, Durbin MK, Budenz DL, et al; Cirrus OCT Normative Database Study Group. Profile and predictors of normal ganglion cell–inner plexiform layer thickness measured with frequency-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2011;52(11):7872-7879.

62. Wang YX, Pan Z, Zhao L, You QS, Xu L, Jonas JB. Retinal nerve fiber layer thickness: the Beijing Eye Study 2011. *PLoS One*. 2013;8(6):e66763.

63. Zhao JJ, Zhuang WJ, Yang XQ, Li SS, Xiang W. Peripapillary retinal nerve fiber layer thickness distribution in Chinese with myopia measured by 3D-optical coherence tomography. *Int J Ophthalmol*. 2013;6(5):626-631.

64. Alasil T, Wang K, Keane PA, et al. Analysis of normal retinal nerve fiber layer thickness by age, sex, and race using spectral domain optical coherence tomography. *J Glaucoma*. 2013;22(7):532-541.

65. Kang SH, Hong SW, Im SK, Lee SH, Ahn MD. Effect of myopia on the thickness of the retinal nerve fiber layer measured by Cirrus HD optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2010;51(8):4075-4083.

66. Weinreb RN. Evaluating the retinal nerve fiber layer in glaucoma with scanning laser polarimetry. *Arch Ophthalmol*. 1999;117(10):1403-1406.

67. Savini G, Barboni P, Parisi V, Carbonelli M. The influence of axial length on retinal nerve fibre layer thickness and optic-disc size measurements by spectral-domain OCT. *Br J Ophthalmol*. 2012;96(1):57-61.

68. Garas A, Tóth M, Vargha P, Holló G. Comparison of repeatability of retinal nerve fiber layer thickness measurement made using the RTVue Fourier-domain optical coherence tomograph and the GDx scanning laser polarimeter with variable or enhanced corneal compensation. *J Glaucoma*. 2010;19(6):412-417.

69. Michelessi M, Lucenteforte E, Oddone F, et al. Optic nerve head and fibre layer imaging for diagnosing glaucoma. *Cochrane Database Syst Rev*. 2015;11(11):CD008803.

70. Choi YJ, Jeoung JW, Park KH, Kim DM. Glaucoma detection ability of ganglion cell–inner plexiform layer thickness by spectral-domain optical coherence tomography in high myopia. *Invest Ophthalmol Vis Sci*. 2013;54(3):2296-2304.

71. Shoji T, Sato H, Ishida M, Takeuchi M, Chihara E. Assessment of glaucomatous changes in subjects with high myopia using spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2011;52(2):1098-1102.

72. Kim NR, Lee ES, Seong GJ, et al. Comparing the ganglion cell complex and retinal nerve fibre layer measurements by Fourier domain OCT to detect glaucoma in high myopia. *Br J Ophthalmol*. 2011;95(8):1115-1121.

73. Jonas JB. Role of cerebrospinal fluid pressure in the pathogenesis of glaucoma. *Acta Ophthalmol*. 2011;89(6):505-514.

74. Siaudvytyte L, Januleviciene I, Daveckaite A, et al. Literature review and meta-analysis of translaminar pressure difference in open-angle glaucoma. *Eye (Lond)*. 2015;29(10):1242-1250.

75. Skarf B. Retinal nerve fibre layer loss in diabetes mellitus without retinopathy. *Br J Ophthalmol*. 2002;86(7):709.

76. van Dijk HW, Verbraak FD, Kok PH, et al. Decreased retinal ganglion cell layer thickness in patients with type 1 diabetes. *Invest Ophthalmol Vis Sci*. 2010;51(7):3660-3665.

77. van Dijk HW, Verbraak FD, Stehouwer M, et al. Association of visual function and ganglion cell layer

thickness in patients with diabetes mellitus type 1 and no or minimal diabetic retinopathy. *Vision Res*. 2011;51(2):224-228.

78. Demir M, Oba E, Sensoz H, Ozdal E. Retinal nerve fiber layer and ganglion cell complex thickness in patients with type 2 diabetes mellitus. *Indian J Ophthalmol*. 2014;62(6):719-720.

79. Chhablani J, Sharma A, Goud A, et al. Neurodegeneration in type 2 diabetes: evidence from spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2015;56(11):6333-6338.

80. Salvi L, Plateroti P, Balducci S, et al. Abnormalities of retinal ganglion cell complex at optical coherence tomography in patients with type 2 diabetes: a sign of diabetic polyneuropathy, not retinopathy. *J Diabetes Complications*. 2016;30(3):469-476.

81. Barber AJ. A new view of diabetic retinopathy: a neurodegenerative disease of the eye. *Prog Neuropsychopharmacol Biol Psychiatry*. 2003;27(2):283-290.

82. Kern TS, Barber AJ. Retinal ganglion cells in diabetes. *J Physiol*. 2008;586(18):4401-4408.

83. Srivastav K, Saxena S, Mahdi AA, Kruzliak P, Khanna VK. Increased serum urea and creatinine levels correlate with decreased retinal nerve fibre layer thickness in diabetic retinopathy. *Biomarkers*. 2015;20(6-7):470-473.

84. Shim SH, Sung KC, Kim JM, et al; Epidemiologic Survey Committee of the Korean Ophthalmological Society. Association between renal function and open-angle glaucoma: the Korea National Health and Nutrition Examination Survey 2010-2011. *Ophthalmology*. 2016;123(9):1981-1988.

85. Savige J, Ratnaike S, Colville D. Retinal abnormalities characteristic of inherited renal disease. *J Am Soc Nephrol*. 2011;22(8):1403-1415.

86. Chamberlain MD, Guymer RH, Dirani M, Hopper JL, Baird PN. Heritability of macular thickness determined by optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2006;47(1):336-340.

87. van Koolwijk LM, Despriet DD, van Duijn CM, et al. Genetic contributions to glaucoma: heritability of intraocular pressure, retinal nerve fiber layer thickness, and optic disc morphology. *Invest Ophthalmol Vis Sci*. 2007;48(8):3669-3676.

88. Hougaard JL, Kessel L, Sander B, Kyvik KO, Sørensen TI, Larsen M. Evaluation of heredity as a determinant of retinal nerve fiber layer thickness as measured by optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2003;44(7):3011-3016.

89. Kuo JZ, Zangwill LM, Medeiros FA, et al. Quantitative trait locus analysis of SIX1-SIX6 with retinal nerve fiber layer thickness in individuals of European descent. *Am J Ophthalmol*. 2015;160(1):123-30.e1.

90. Carnes MU, Liu YP, Allingham RR, et al; NEIGHBORHOOD Consortium Investigators. Discovery and functional annotation of *SIX6* variants in primary open-angle glaucoma. *PLoS Genet*. 2014;10(5):e1004372.

91. Cheng CY, Allingham RR, Aung T, et al. Association of common *SIX6* polymorphisms with peripapillary retinal nerve fiber layer thickness: the Singapore Chinese Eye Study. *Invest Ophthalmol Vis Sci*. 2014;56(1):478-483.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

The correlation between cognitive performance and retinal nerve fibre layer thickness is largely explained by genetic factors

Received: 21 July 2016
Accepted: 07 September 2016
Published: 28 September 2016

Eneh Jones-Odeh[1,*], Ekaterina Yonova-Doing[1,*], Edward Bloch[1], Katie M. Williams[2], Claire J. Steves[1] & Christopher J. Hammond[1,2]

Retinal nerve fibre layer (RNFL) thickness has been associated with cognitive function but it is unclear whether RNFL thinning is secondary to cortical loss, or if the same disease process affects both. We explored whether there is phenotypic sharing between RNFL thickness and cognitive traits, and whether such sharing is due to genetic factors. Detailed eye and cognitive examination were performed on 1602 twins (mean age: 56.4 years; range: 18–89) from the TwinsUK cohort. Associations between RNFL thickness and ophthalmic, cognitive and other predictors were assessed using linear regression or analysis of variance models. Heritability analyses were performed using uni- and bivariate Cholesky decomposition models. RNFL was thinner with increase in myopia and with decrease in disc area ($p < 0.001$). A thicker RNFL was associated with better performance on mini mental state examination (MMSE, $F(5,883) = 5.8$, $p < 0.001$), and with faster reaction time (RT, $\beta = -0.01$; $p = 0.01$); independent of the effects of age, refractive error and disc area ($p < 0.05$). RNFL thickness was highly heritable (82%) but there was low phenotypic sharing between RNFL thickness and MMSE (5%, 95% CI: 0–10%) or RT (7%, 95% CI: 1–12%). This sharing, however, was mostly due to additive genetic effects (67% and 92% of the shared variance respectively).

Cognitive function has been defined as the ability to appropriately identify and manage complex external or internal stimuli, and encompasses many mental abilities and processes, including decision making, memory, attention, and problem solving[1]. The retinal nerve fibre layer (RNFL) has been shown to be thinner in people with memory loss and cognitive decline in certain neurodegenerative conditions such as Alzheimer's disease (AD) and multiple sclerosis (MS)[2]. The retina forms as a functional extension of the central nervous system (CNS), developing as a result of diencephalic evagination of pluripotent cells during embryogenesis[3] and shares similar morphological properties with other neurons in the CNS[4]. These properties allow the opportunities for studying the retina as a model of the CNS, and provide window for accessible and direct observation of the CNS using ocular imaging techniques such as optical coherence tomography (OCT).

OCT is now the most commonly used optical imaging technique for structural assessment of the retina, allowing for accessible, inexpensive and non-invasive *in vivo* quantitative measurement of the optic nerve head (ONH) and RNFL. With similar principles to that of ultrasound, OCT uses low-coherence interferometry to measure optical backscattering of light from the retina and a reference mirror and calculates RNFL thickness (RNFLT) from generated two-dimensional data sets.

At a population level, a thinner RNFLT has been associated with lower cognitive performance[5–7] in cross-sectional analysis, although the Erasmus Rucphen Family (ERF) Study only found this association in younger individuals[6] and the association was reversed in the older Lothian Birth Cohort Study when corrected for IQ at age 11[5]. It is also well-described in populations with mild cognitive impairment[2]. However it is unclear whether RNFL thinning is secondary to cortical loss (retrograde damage), or if the same disease process affects both brain and retina.

[1]Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom. [2]Ophthalmology, King's College London, London, United Kingdom. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.J.H. (email: chris.hammond@kcl.ac.uk)

	All subjects (N = 1602)		MZ twins (N = 947)		DZ twins (N = 655)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age(years)*	56.4 (15.2)	18.2–89.8	54.5 (16.0)	18.2–88.8	59.1 (13.4)	19.1–89.8
Average RNFL (μm)	96.8 (9.2)	66.6–123.2	97.1 (9.1)	70.1–123.2	96.3 (9.3)	66.6–120.3
Superior (μm)	118.2 (13.4)	74.7–158.9	118.3 (13.4)	74.4–155.5	118.1 (13.5)	74.7–158.9
Nasal (μm)	75.1 (10.8)	40.0–112.5	75.3 (10.9)	40.0–112.5	74.7 (10.7)	46.0–110.9
Inferior (μm)	121.4 (13.7)	73.3–167.5	122.0 (13.6)	76.4–167.5	120.5 (13.8)	73.3–159.7
Temporal (μm)	72.4 (8.7)	43.6–102.3	72.8 (8.9)	43.6–102.3	71.8 (8.4)	48.5–101.1
SphE (D)	−0.1 (0.1)	−11.3–9.3	−0.2 (2.4)	−11.3–7.5	0.1 (2.3)	−9.3–9.3
IOP (mmHg)	13.4 (2.7)	6.7–23.8	13.4 (2.7)	7.0–23.8	13.5 (2.7)	6.6–22.5
Corneal Thickness (μm)	538 (38.3)	221.2–752.3	537.7 (38.3)	221.3–752.3	538.4 (38.4)	343.0–745.6
Reaction time (ms)*	586.7 (87.9)	341.1–928.3	566.7 (101.7)	341.1–967.9	583.7 (95.5)	368.7–1049.5
	%		%		%	
Gender						
Female/Male	88.6/11.4		87.9/12.1		90.9/9.1	
Education*						
Primary	19.1		16.4		22.4	
Secondary	53.4		51.5		55.7	
Higher	27.5		32.0		21.9	
IMD						
Q1	6.9		7.4		6.2	
Q2	12.0		11.8		12.2	
Q3	18.6		19.1		18.1	
Q4	26.9		27.5		26.3	
Q5	35.7		34.3		37.3	
MMSE score*						
≤25	1.3		0.8		2.1	
26	2.1		2.3		1.9	
27	5.0		4.0		6.5	
28	12.5		11.7		13.5	
29	29.8		27.9		32.6	
30	49.2		53.2		43.3	
Recall*						
≤1	5.4		4.1		3.0	
2	21.9		20.3		10.0	
3	72.6		75.6		68.4	
Verbal fluency						
<11	16.4		16.6		16.0	
≥11	83.6		83.4		84.0	

Table 1. Characteristics of the study population. Legend: This table shows summary data for the different variables studied for the whole sample population and by zygosity status. The *denotes variables which were significantly different ($p < 0.05$) between monozygotic (MZ) and dizygotic (DZ) twins. RNFL –retinal nerve fibre layer; μm – micrometers; SphE – spherical equivalent; IOP – Intraocular pressure. ms – milliseconds, IMD – index of multiple deprivation from Q1 – bottom 20% most deprived to Q5 – top 20% most affluent.

A twin and a family study have demonstrated heritability of the RNFLT[8,9], but to date, to our knowledge, no studies have explored whether this phenotypic sharing is due to shared genetic or environmental factors.

The aim of this large twin study was to examine whether the RNFLT measured by OCT was associated with cognitive function and whether this association was driven by shared genetic factors. A further objective of this study was to explore associations between RNFLT and age, sex, spherical equivalent (SphE), and other factors known to influence glaucoma such as intraocular pressure (IOP) and optic disc size.

Results

Association analysis. We initially explored associations between RNFL thickness and various ophthalmological and cognitive measures (Table 1). 1602 twins of 908 families (mean age 56.4 years; range 18–89) from the TwinsUK cohort were included (Fig. 1). There was no difference between monozygotic (MZ) and dizygotic (DZ) twins except for the following variables ($p < 0.05$): age, reaction time (RT), MMSE score and recall (RC) (Table 1). The MZ twins were on average slightly younger than DZ twins (55 versus 59 years of age) and had better reaction times, MMSE score and recall. When adjusted for age, however, the three cognitive parameters did not differ between MZ and DZ twins. The inferior quadrant had the highest RNFL thickness followed by superior, nasal and

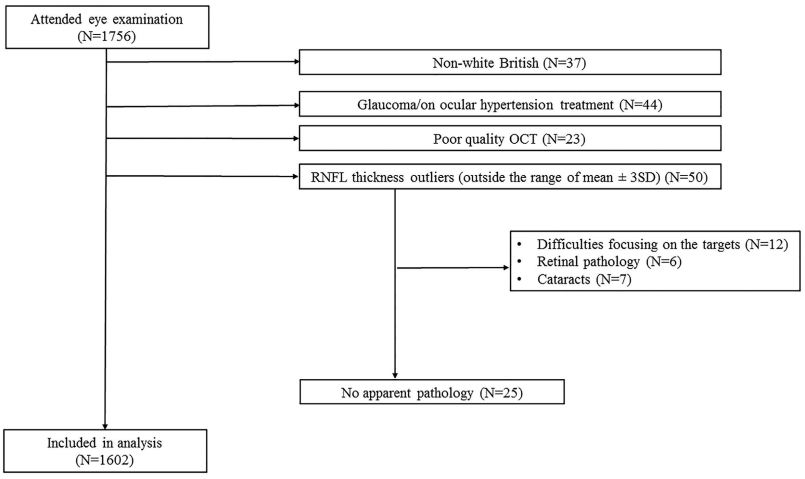


Figure 1. Consort diagram of the study. Legend: This figure presents a consort diagram of the study showing reasons for exclusion from the study.

temporal quadrants. The differences in thickness between quadrants were statistically significant ($p < 1 \times 10^{-4}$). All of the RNFL parameters were approximately normally distributed.

Table 2 summarises the univariable (A) and multivariable (B) results from association analysis for the quantitative predictors. Increase in age was associated with decrease in RNFL thickness for the average of the four quadrants and in each quadrant individually ($p < 0.001$). Education was nominally associated with average RNFL thickness and with the inferior and temporal quadrants in the univariable analysis only ($p < 0.05$). Increase in SphE (more hyperopic refraction) and in disc area (larger optic discs) were strongly associated ($p < 0.001$) with increase in RNFL thickness for the average of the four quadrants and in each quadrant, except for SpheE in the temporal quadrant ($p = 0.41$). Disc area remained strongly associated ($p < 0.001$) in the multivariable analysis with the average of the four quadrants and in all the quadrants separately, while age and SphE were associated with RNFLT for all but the temporal quadrant (Table 2). There was no association between sex, index of multiple deprivation (IMD) or IOP corrected for central corneal thickness (CCT) and RNFLT (Table 2).

None of the cognitive parameters were associated with SphE, education or IMD, independently of age ($p < 0.05$). A faster reaction time was associated with a thinner RNFL thickness ($p < 0.001$) and this stayed significant in the multivariable model for average RNFL ($p = 0.01$), and for the inferior and temporal quadrants ($p = 0.01$ and $p = 0.04$ respectively). As cataract surgery has been reported to improve reaction times[10], we compared the reaction time in the 63 pseudophakic participants to that of 126 age-matched individuals without cataract surgery (case:control ratio of 1:2). There was no difference between the two groups (mean of 630 ms and 632 ms respectively, $p = 0.8$). Cataract surgery status was associated with RNFL thickness in univariable analysis ($\beta = -3.29$, $p = 0.02$), but not in the multivariable analysis (likely to due to the association between cataract surgery and age).

There were significant differences in RNFL thickness between MMSE groups ($p_{trend} < 0.003$). The one-way ANOVA results for MMSE were as follows: average RNFL – $F(5, 883) = 5.8$, $p < 0.001$; Superior quadrant – $F(5, 883) = 4.0$, $p = 0.001$; Inferior quadrant – $F(5, 883) = 5.9$, $p < 0.001$; Temporal quadrant – $F(5, 883) = 2.5$, $p = 0.02$; Nasal quadrant – $F(5,883) = 2.8$, $p = 0.02$. These differences were still significant after taking age, SphE and disc area into account, as determined by MANOVA for all but the temporal quadrant: average RNFL – $F(5,615) = 4.2$, $p = 0.001$; Superior quadrant – $F(5,615) = 3.1$, $p = 0.01$; Inferior quadrant – $F(5,615) = 3.8$, $p = 0.002$; Nasal quadrant – $F(5,474) = 2.6$, $p = 0.02$. Recall was associated with RNFL thickness only in the inferior quadrant $-F(5,886) = 5.2$, $p = 0.01$, however this effect was not independent from the effects of age, SphE and disc area $F(2,618) = 2.9$, $p = 0.06$. No statistically significant association was detected between RNFL thickness and verbal fluency (VF) (data not shown).

Heritability analysis. The results of the univariate heritability models are reported in Table 3. In all cases the best fitting model was the AE (addictive genetic effects/unique environment) model. As can be seen from the table, RNFL thickness is highly heritable with additive genetic effects accounting for 65% (Nasal quadrant) to 83% (the average of the four quadrants) of variance. As pointed above, due to the fact that the MZ twins were younger than the DZ twins, there were differences between the two groups in terms of MMSE, RC and RT. For this reason we adjusted those predictors for the effect of age prior to the heritability analysis. The measurements of cognitive function were moderately heritable (MMSE = 32%, RT = 44% and VF = 48%), with the exception of recall which showed low heritability (19%).

	(A) Univariable														
	Average RNFL			Superior quadrant			Nasal quadrant			Inferior quadrant			Temporal quadrant		
	beta	se	p	beta	se	p	beta	se	p	beta	se	p	beta	se	p
Age	−0.15	0.02	<0.001	−0.21	0.03	<0.001	−0.13	0.02	<0.001	−0.19	0.03	<0.001	−0.07	0.02	<0.001
Sex	0.46	1.56	0.76	0.32	2.17	0.88	−0.85	1.89	0.65	1.61	2.48	0.52	0.77	1.36	0.57
Education	0.97	0.46	0.04	1.06	0.68	0.12	0.43	0.54	0.43	1.33	0.65	0.043	1.05	0.42	0.01
IMD	0.19	0.24	0.42	−0.05	0.35	0.90	0.25	0.28	0.37	0.24	0.34	0.47	0.32	0.22	0.14
SphE	1.14	0.13	<0.001	1.49	0.19	<0.001	1.43	0.15	<0.001	1.54	0.21	<0.001	0.12	0.15	0.41
IOP	−0.10	0.15	0.50	−0.24	0.22	0.13	0.10	0.17	0.61	−0.06	0.21	0.77	−0.22	0.13	0.09
IOPc	−0.06	0.11	0.59	−0.01	0.16	0.93	0.02	0.12	0.85	−0.15	0.16	0.33	−0.09	0.10	0.36
Disc area	8.53	0.79	<0.001	10.16	1.17	<0.001	8.32	1.00	<0.001	11.93	1.54	<0.001	3.68	0.82	<0.001
Reaction time	−0.02	0.003	<0.001	−0.02	0.004	<0.001	−0.01	0.003	<0.001	−0.02	0.004	<0.001	−0.01	0.003	<0.001
	(B) Multivariable														
	Average RNFL			Superior			Nasal			Inferior			Temporal		
	beta	se	p	beta	se	p	beta	se	p	beta	se	p	beta	se	p
Age	−0.16	0.02	<0.001	−0.27	0.03	<0.001	−0.19	0.02	<0.001	−0.20	0.04	<0.001	—	—	—
Education	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SphE	1.18	0.13	<0.001	1.43	0.22	<0.001	1.33	0.16	<0.001	1.84	0.19	<0.001	—	—	—
Disc area	6.03	1.01	<0.001	6.43	1.19	<0.001	5.39	0.98	<0.001	7.70	1.20	<0.001	3.59	0.85	<0.001
Reaction time	−0.01	0.004	0.01	—	—	—	—	—	—	−0.01	0.01	0.01	−0.01	0.003	0.002

Table 2. Univariable and Multivariable Linear Regression Analyses of RNFL thickness and ophthalmic, wcognitive and other predators. Legend: This table shows the results from the univariable (**A**) and multivariable (**B**) analysis. RNFL – retinal nerve fibre layer; IMD – Index of multiple deprivation; SphE – spherical equivalent; IOP – Intraocular pressure; IOPc – IOP corrected for corneal thickness by adding thickness as predictor in the regression; se – standard error, p – p value. Results for the variable with p-value > 0.05 in the stepwise regression are not reported.

	average RNFL		Superior quadrant		Nasal quadrant		Inferior quadrant		Temporal quadrant	
	estimate	95% CI	estimate	95% CI	estimate	95% CI	estimate	95% CI	estimate	95% CI
A	0.83	0.66–0.85	0.74	0.58–0.77	0.65	0.60–0.69	0.75	0.72–0.79	0.74	0.63–0.78
C	0.00	0.0–0.0	0.00	0.0–0.0	0.00	0.0–0.0	0.00	0.0–0.0	0.00	0.0–0.0
E	0.17	0.15–0.20	0.26	0.23–0.31	0.35	0.31–0.40	0.25	0.21–0.29	0.26	0.22–0.30
	MMSE score*		Recall*		Reaction time*		Verbal fluency			
	estimate	95% CI	estimate	95% CI	estimate	95% CI	estimate	95% CI		
A	0.32	0.23–0.40	0.19	0.09–0.28	0.44	0.37–0.51	0.48	0.41–0.54		
C	0.00	0.0–0.0	0.00	0.0–0.0	0.00	0.0–0.0	0.00	0.0–0.0		
E	0.68	0.60–0.77	0.81	0.79–0.91	0.56	0.49–0.63	0.52	0.46–0.59		

Table 3. Univariate Heritability estimates of RNFL thickness and cognitive variables. Legend: This table presents the results from univariate heritability modelling for the different RNFL and cognition parameters. RNFL –retinal nerve fibre layer; A –additive genetic factors; C –common/shared environmental factors; E –unique environmental latent factors; MMSE – mini-mental state examination. *Scores were adjusted for age prior to the heritability analysis.

When taking into account the effect of age, RNFL thickness shared 5% (95% CI: 0−10%) of phenotypic variance with MMSE and 7% with RT (95% CI: 1–12%) (Fig. 2). Additive genetic effects accounted for 67% (95% CI: 61–81%) and 87% (95% CI: 82–99%) of the shared variance for MMSE and RT respectively and rest was explained by environmental factors. Adjusting for SphE and disc area did not substantially change the results: phenotypic sharing between MMSE and RNFL was 8% (95% CI: 2–15%) and of those 93% (72–96%) was due to additive genetic factors; phenotypic sharing between RT and RNFL was 11% (95% CI: 4–7%) and of those 82% (75–94%) was due to additive genetic factors.

Discussion

This study, based on an unselected twin population, has confirmed an association between thinner RNFL and lower cognitive scores, and suggests this phenotypic sharing is largely due to shared genetic factors. This raises the possibility that the processes involved in cognitive decline and ultimately dementia are similar to those resulting in loss of RNFL, and the latter could therefore be a biomarker of dementia. However, the degree of phenotypic sharing (8% for MMSE score and 11% for reaction time, when adjusted for age, refractive error and optic disc size) is relatively small, and given other factors such as glaucoma influence RNFL thickness, it remains to be seen if this will have enough specificity and sensitivity to act as a biomarker of disease progression in dementia.

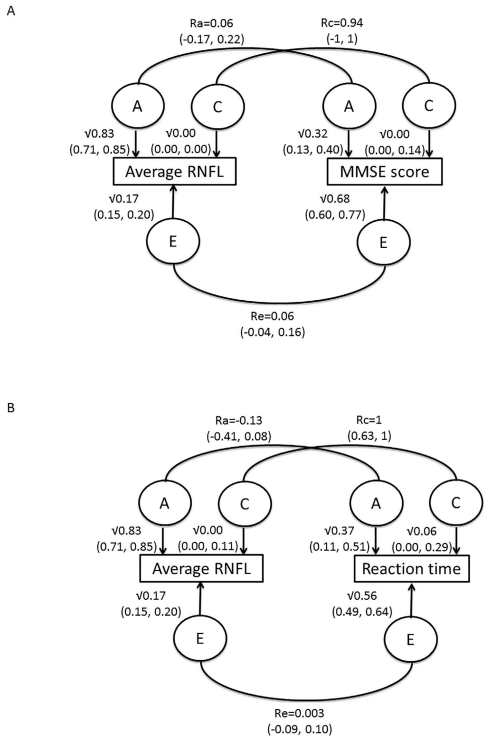


Figure 2. Bivariate heritability model showing the amount of shared genetic variation for average RNFL thickness and MMSE (**A**) and Reaction time (**B**). Legend: This figure presents the bivariate heritability models for RNFL thickness and the associated cognitive parameters. All traits were adjusted for age prior to the analysis. RNFL – retinal nerve fibre layer; MMSE – mini-mental state examination; A – additive genetic factors; C – common/shared environmental factors; E – unique environmental latent factors; Ra – genetic correlation; Rc – common environment correlation; Re – unique environment correlation.

Our findings are similar to another UK-based aging cohort from the Norfolk EPIC study by Khawaja *et al.*[7], which showed significant, albeit weaker, associations with MMSE and also the Hopkins Verbal Learning Test (testing recognition, learning and memory) and the National Adult Reading Test (NART, testing premorbid intelligence). Similarly, the Dutch ERF study[6] showed better cognitive function associated with thicker RNFL across various measures, with 2.8% of cognitive scores explained by RNFLT, although somewhat surprisingly the strongest effects seemed to be in the younger age groups (below 40 years). Some of these differences may be due to measurement techniques, and these studies used HRT (Heidelberg Retina Tomograph) and scanning laser polarimetry (the GDx VCC) rather than the current gold standard used in this study, of OCT. The Lothian study[5] using OCT in only 96 subjects aged 73 years, found that while cognitive decline seemed to be associated with thinner RNFL, when corrected for IQ at age 11, those with thicker RNFL had worse cognition, again somewhat counterintuitively. We do not have historic IQ data on the TwinsUK cohort, but did not observe any relationship between the educational and socio-economic status and either RNFL or the cognitive parameters in this study. The cognitive measures used in each study differed, which may also explain some of the differences; we selected MMSE, RT and VF because these tests are commonly used in clinical settings[11–14].

Our results showed that the average RNFL was significantly thinner with increasing age, lower spherical equivalent (i.e. myopia) and with smaller optic disc size. Similar findings have been confirmed in other studies[15,16] and an inverse relationship between RNFLT and SphE identified in a small twin study of 50 pairs[8]. We found the phenotypic sharing between RNFL and cognition was greater when these factors were accounted for, suggesting future studies examining RNFLT variation should measure refractive error or axial length, which is was measured in the Khawaja *et al.* study[7] but is not commonly done in non-ophthalmic studies, in addition to adjusting for age and disc size (which is measured by OCT).

Our community-based study confirms the relationship between RNFL and cognitive decline in clinic-based studies which found a thinner RNFL in patients even with mild cognitive impairment[17–20]. In particular, the inferior quadrant has been proposed to be more susceptible to neuronal loss[2,14,21,22]. Certainly in our data, and others[5], the inferior quadrant showed the strongest relationships with cognitive tests, although as the quadrant with the thickest RNFL it is unsurprising that the effect size is largest in ANOVA analyses, but despite this, the inferior (along with the temporal) quadrant was the only quadrant which remained significantly associated with reaction time when age, spherical equivalent and disc size were included in the multivariable model. Histological studies have shown significant correlation between retinal layers determined by OCT and retinal histology[23] demonstrated reduced RNFL thickness in patients with AD when compared with normal controls[24], however individual quadrants of the RNFL were not examined separately. It is unclear why the inferior RNFL might be preferentially affected but it may be due to the fact that nerve fibers with larger diameter atrophy more rapidly and those fibers are more abundant in inferior and superior quadrants[25].

RNFL thickness was shown in this large twin study of 730 pairs to be highly heritable, while MMSE score and RT were moderately heritable with additive genetic effects accounting for 82%, 32% and 44% of variance respectively. These results are in line with previous twin studies[8,26], and as expected, higher than found using other family-based study designs[9]. Many quantitative traits that can be measured accurately are highly heritable, such as refractive error and height, and both these traits appear to be influenced by hundreds of common genetic variants[27–29]. Genome-wide association studies (GWAS) of RNFL are in progress, and we await with interest whether genes or pathways identified in these analyses are common to dementia GWAS.

This study is not without its limitations. First, as the TwinsUK cohort consists mainly of white British middle-aged females (mean 56.2 years, SD 15.4) we were not able to look at gender or ethnicity differences. However, our baseline measurements of average RNFL thickness ($96.2 \pm 9.1\,\mu m$) are similar to those obtained in healthy controls of other studies[11,12,30] that have investigated the cognitive associations of RNFL thickness using OCT ($91.5 \pm 7.1\,\mu m$), ($98.6 \pm 1.67\,\mu m$), ($94.3 \pm 11.3\,\mu m$) respectively, and quadrant thickness distribution (ISNT rule: inferior RNFL thickest, followed by superior, nasal and temporal) comparable to those obtained in similar studies[5,14,31]. We found no differences between genders for RNFL (and with 183 males the sample size in not insignificant). We therefore believe that these data are generalisable to the wider European population. Second, we do not have longitudinal cognitive or RNFL data on these subjects, so causality cannot be applied to any of the cross-sectional analyses, and these results may still be due to residual confounding by age, or refractive error, not fully corrected in linear models, or other confounders. Third, due to the low phenotypic sharing we observed, the bivariate heritability results were of borderline statistical significance, and larger studies might be needed for further analyses, including examining the amount of variance explained by known genetic variants from dementia genome-wide association studies.

In conclusion, we have shown RNFL thickness to be associated with MMSE and reaction time cognitive performance in a healthy adult population, and the phenotypic sharing was largely due to shared genetic factors. However, given the relatively low phenotypic sharing, RNFL thickness is probably a poor predictor of cognitive decline in an unselected population, although it remains to be determined in longitudinal studies, combining several and different cognitive tests, how predictive RNFL thickness might be. Further genetic studies may provide insight into shared pathophysiology between loss of RNFL and cognitive impairment.

Methods

Subjects. The TwinsUK cohort is a volunteer twin registry of 11,000 British adults, recruited from the general population through national media campaigns in the United Kingdom, and representative to the broader population in terms of disease-related and lifestyle characteristics[32]. All subjects were unaware of the hypotheses in this study at the time of enrolment. The studies in the TwinsUK cohort were performed with the approval of the Guys and ST Thomas' Ethics Committee, and participants signed a written informed consent in accordance with the Declaration of Helsinki. All methods were performed in accordance with the relevant guidelines and regulations.

Between August 2014 and March 2016, 1756 twins from 974 families underwent detailed eye examination which included RNFL thickness measurements using spectral domain optical coherence tomography (Optovue iVue system, Fermont, California), non-cycloplegic autorefraction, non-contact IOP and central corneal thickness (CCT) measurements (Visionix120, The Luneau Technology group) of both eyes. Of the 1756 individuals, 154 were excluded, leaving 1602 twins of 908 families in the final analysis (Fig. 1). The reasons for exclusion were as follows: 37 individuals were of non-white British ethnicity; 44 individuals had self-reported diagnoses of glaucoma or were on treatment for ocular hypertension; 23 had poor quality OCT scans; and 50 individuals were outliers (fell outside the range of mean $\pm$ 3 SD) in terms of RNFL thickness (either average RNFL or in one of the 4 quadrants). For 25 out of the 50 outliers there was no obvious explanation why their measurements were so different to the mean and the scans for those individuals were deemed of good quality by an ophthalmologist. Of the remaining outliers, 12 individuals were described as having difficulty with target fixation, 6 presented with retinal pathology and 7 reported having cataracts in the eye/eyes with the aberrant RNFL measurement. From the 1602 twins included in the study, 63 (3.9%) reported having had cataract surgery previously.

In addition, data on age, sex, education (primary, secondary and higher) and socio-economic status (Index of Multiple Deprivation - IMD) were also collected.

Cognitive assessment. As part of the visit, the twins also underwent cognitive assessment including the standardised mini-mental state examination (MMSE), verbal fluency (VF), and reaction time (RT). Recall (RC), a subtest included in the MMSE was extrapolated as an individual test and results obtained from this subtest were analysed separately. Cognitive data was available for 1593 individuals for the VF, MMSE and recall tests, and 1563 individuals for the RT test.

The MMSE is used widely in clinical settings as a standard screening tool for global measurement of cognitive function[33]. It contains 20 subtests with a maximum score of 30 and helps facilitate detection of cognitive impairment or change in mental status[34]. We decided to analyse RC scores separately as it has been shown to provide a good measure of episodic memory, and some other components of the full MMSE may not adequately discriminate healthy individuals in a population with mild cognitive impairment[35].

The verbal fluency tasks assesses the speed and ease at which individuals can access words from memory[36], and performance has been associated with both temporal and frontal lobe activity[37,38]. VF was tested by examiners instructing participants to name as many words beginning with a given letter of the alphabet, excluding names of people or places, and a further test required participants to name as many animals beginning with any letter. Participants were given 1 minute for each component of the VF test and were marked out of 7 for each task, where 0 = 2 answers or less and 7 = 17 answers or more. The scores from the two tasks were summed and individuals were categorised into two groups (normal performance, low performance) with cut-off score of 11.

RT was measured using the 'Deary-Liewald Reaction Time Tester' on a Windows 7 computer and was measured in milliseconds[39]. The test comprises of two components, the simple reaction time test (pressing a key anytime a target appears) and the choice reaction time test (pressing a key corresponding to 1 of 4 boxes where a target appears). Participants had 8 practice trials and 20 experimental trials for each component of the test. In the present study we analysed the mean of the two tests.

Statistical Analysis. *Association analysis.* Comparisons of means and proportions for all variables between Monozygotic (MZ) and Dizygotic (DZ) twins or between RNFL quadrants were performed using two-sample two-tailed t-tests or z-tests, assuming equal variance. Associations between RNFL (either average or per quadrant) and age, sex, SphE, IOP and reaction time was assessed using univariable linear regression analyses, followed by multivariable linear regression model for factors showing significant ($p < 0.05$) association in the univariable model. Independent variables were identified using stepwise backwards procedure with threshold for removal set at 0.05. In all regression models family structure was taken into account.

Associations between RNFL thickness (either average or per quadrant) and MMSE, verbal fluency and recall were assessed using one-way ANOVA, followed by MANOVA that included age, SphE and disc area into the model. For these analyses only one twin per pair was selected at random. As there was strong correlation between the two eyes, we used the mean measurements of the two eyes for all eye phenotypes. All analyses were carried using STATA14 statistical package (www.stata.com).

Modelling of Heritability. Heritability analyses were performed on 730 twin pairs (438 monozygotic and 292 dizygotic) as data was missing on 178 co-twins. Zygosity was determined by a standardised questionnaire and confirmed using genome-wide single nucleotide polymorphism genotyping data or DNA short tandem repeat fingerprinting.

Univariate and bivariate Cholesky decomposition heritability models of RNFL thickness and the associated cognitive phenotypes were calculated using maximum likelihood structural equation twin modelling implemented in the OpenMx package (http://openmx.psyc.virginia.edu). The univariate method separates phenotypic variation in additive (A) genetic factors and common/shared (C) or unique environmental (E) latent factors, while bivariate model defines the amount of shared genetic variation that explains a given amount of shared phenotypic variance between two or more traits. In cases where the intraclass correlation coefficient for MZ twins was more than twice that of the DZ twins, ADE models were also computed (D stands for Dominant factors). The goodness of fit of the full and reduced ACE/ADE models were compared with the observed data and the best fitting model was selected. Two different bivariate models were explored. In Model 1 RNFL thickness was adjusted for age prior to the analysis, while in Model 2 RNFL was also adjusted for SphE and disc area in addition to age. In both models the cognitive traits were adjusted for age only.

References

1. Blazer, D. G. *Cognitive Aging: Progress in Understanding and Opportunities for Action*. 17–30 (National Academies Press, 2015).
2. Jones-Odeh, E. & Hammond, C. J. How strong is the relationship between glaucoma, the retinal nerve fibre layer, and neurodegenerative diseases such as Alzheimer's disease and multiple sclerosis? *Eye* **29,** 1270–1284 (2015).
3. Cvekl, A. & Tamm, E. R. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. *Bioessays* **26,** 374–386 (2004).
4. London, A., Benhar, I. & Schwartz, M. The retina as a window to the brain—from eye research to CNS disorders. *Nat. Rev. Neurol.* **9,** 44–53 (2013).
5. Laude, A. *et al.* Retinal nerve fiber layer thickness and cognitive ability in older people: the Lothian Birth Cohort 1936 study. *BMC Ophthalmol.* **13,** 28 (2013).
6. van Koolwijk, L. M. E. *et al.* Association of Cognitive Functioning with Retinal Nerve Fiber Layer Thickness. *Investig. Opthalmology Vis. Sci.* **50,** 4576 (2009).
7. Khawaja, A. P. *et al.* Retinal Nerve Fiber Layer Measures and Cognitive Function in the EPIC-Norfolk Cohort Study. *Invest. Ophthalmol. Vis. Sci.* **57,** 1921–1926 (2016).
8. Hougaard, J. L. *et al.* Evaluation of heredity as a determinant of retinal nerve fiber layer thickness as measured by optical coherence tomography. *Invest. Ophthalmol. Vis. Sci.* **44,** 3011–3016 (2003).
9. van Koolwijk, L. M. E. *et al.* Genetic Contributions to Glaucoma: Heritability of Intraocular Pressure, Retinal Nerve Fiber Layer Thickness, and Optic Disc Morphology. *Investig. Opthalmology Vis. Sci.* **48,** 3669 (2007).
10. Schmoll C. *et al.* Reaction time as a measure of enhanced blue-light mediated cognitive function following cataract surgery. *Br J Ophthalmol.* **95,** 1656–9 (2011).
11. Gao, L., Liu, Y., Li, X., Bai, Q. & Liu, P. Abnormal retinal nerve fiber layer thickness and macula lutea in patients with mild cognitive impairment and Alzheimer's disease. *Arch. Gerontol. Geriatr.* **60,** 162–167 (2015).
12. Kesler, A., Vakhapova, V., Korczyn, A. D., Naftaliev, E. & Neudorfer, M. Retinal thickness in patients with mild cognitive impairment and Alzheimer's disease. *Clin. Neurol. Neurosurg.* **113,** 523–526 (2011).
13. Paquet, C. *et al.* Abnormal retinal thickness in patients with mild cognitive impairment and Alzheimer's disease. *Neurosci. Lett.* **420,** 97–99 (2007).

14. Iseri, P. K., Altinaş, O., Tokay, T. & Yüksel, N. Relationship between Cognitive Impairment and Retinal Morphological and Visual Functional Abnormalities in Alzheimer Disease. *J. Neuroophthalmol.* **26,** 18–24 (2006).
15. Zhao, L., Wang, Y., Chen, C. X., Xu, L. & Jonas, J. B. Retinal nerve fibre layer thickness measured by Spectralis spectral-domain optical coherence tomography: The Beijing Eye Study. *Acta Ophthalmol. (Copenh.)* **92,** e35–41 (2014).
16. Mok, K. H., Lee, V. W.-H. & So, K. F. Retinal nerve fiber layer measurement of the Hong Kong chinese population by optical coherence tomography. *J. Glaucoma* **11,** 481–483 (2002).
17. Liu, D. *et al.* Thinner changes of the retinal nerve fiber layer in patients with mild cognitive impairment and Alzheimer's disease. *BMC Neurol.* **15,** 14 (2015).
18. Shi, Z. *et al.* Greater attenuation of retinal nerve fiber layer thickness in Alzheimer's disease patients. *J. Alzheimers Dis.* **40,** 277–283 (2014).
19. Parisi, V. *et al.* Morphological and functional retinal impairment in Alzheimer's disease patients. *Clin. Neurophysiol.* **112,** 1860–1867 (2001).
20. Tatham, A. J. *et al.* Glaucomatous retinal nerve fiber layer thickness loss is associated with slower reaction times under a divided attention task. *Am. J. Ophthalmol.* **158,** 1008–1017 (2014).
21. Shen, Y. *et al.* Retinal nerve fiber layer thickness is associated with episodic memory deficit in mild cognitive impairment patients. *Curr Alzheimer Res.* **11,** 259–266 (2014).
22. Shi, Z. *et al.* The Utilization of Retinal Nerve Fiber Layer Thickness to Predict Cognitive Deterioration. *J Alzheimers Dis JAD.* **49,** 399–405 (2015).
23. Toth, C. A. *et al.* A comparison of retinal morphology viewed by optical coherence tomography and by light microscopy. *Arch. Ophthalmol. Chic. Ill 1960* **115,** 1425–1428 (1997).
24. Hinton, D. R., Sadun, A. A., Blanks, J. C. & Miller, C. A. Optic-nerve degeneration in Alzheimer's disease. *N. Engl. J. Med.* **315,** 485–487 (1986).
25. Quigley, H. A., Sanchez, R. M., Dunkelberger, G. R., L'Hernault, N. L. & Baginski, T. A. Chronic glaucoma selectively damages large optic nerve fibers. *Invest. Ophthalmol. Vis. Sci.* **28,** 913–920 (1987).
26. Singer, J. J., MacGregor, A. J., Cherkas, L. F. & Spector, T. D. Genetic influences on cognitive function using the Cambridge Neuropsychological Test Automated Battery. *Intelligence* **34,** 421–428 (2006).
27. Guggenheim, J. A. *et al.* Coordinated genetic scaling of the human eye: shared determination of axial eye length and corneal curvature. *Invest. Ophthalmol. Vis. Sci.* **54,** 1715–1721 (2013).
28. Verhoeven, V. J. M. *et al.* Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat. Genet.* **45,** 314–318 (2013).
29. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014).
30. Oktem, E. O. *et al.* The relationship between the degree of cognitive impairment and retinal nerve fiber layer thickness. *Neurol. Sci.* **36,** 1141–1146 (2015).
31. Bendschneider, D. *et al.* Retinal Nerve Fiber Layer Thickness in Normals Measured by Spectral Domain OCT. *J. Glaucoma* **19,** 475–482 (2010).
32. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* dyr207 (2012).
33. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12,** 189–198 (1975).
34. Cummings, J. L. Mini-Mental State Examination: norms, normals, and numbers. *Jama* **269,** 2420–2421 (1993).
35. Carcaillon, L., Amieva, H., Auriacombe, S., Helmer, C. & Dartigues, J.-F. A subtest of the MMSE as a valid test of episodic memory? Comparison with the Free and Cued Reminding Test. *Dement. Geriatr. Cogn. Disord.* **27,** 429–438 (2009).
36. Lezak, M. D. *Neuropsychological Assessment*. (Oxford University Press, 2004).
37. Newcombe, F. *Missile wounds of the brain: A study of psychological deficits*. (1969).
38. Coslett, H. B., Bowers, D., Verfaellie, M. & Heilman, K. M. Frontal verbal amnesia. Phonological amnesia. *Arch. Neurol.* **48,** 949–955 (1991).
39. Deary, I. J., Liewald, D. & Nissan, J. A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary-Liewald reaction time task. *Behav. Res. Methods* **43,** 258–268 (2011).

Acknowledgements
The authors thank all the twin volunteers for their participation. This study was funded by the Wellcome Trust and the International Glaucoma Association reference: 2013 research award. The Twins UK study was also funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study also received support from the National Institute for Health Research–funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. E.Y.-D.: Funding–Biotechnology and Biological Sciences Research Council and the London Interdisciplinary Doctoral Program. K.M.W.: Support–Medical Research Council Clinical Research Training Fellowship. The sponsors or funding organizations had no role in the design or conduct of this research.

Author Contributions
Conception and design: E.J.-O., E.Y.-D., C.S. and C.J.H. Data collection: E.J.-O., E.Y.-D., E.B. and K.M.W. Analysis and interpretation: E.J.-O., E.Y.-D., E.B. and K.M.W. Overall responsibility: E.J.-O., E.Y.-D., C.S. and C.J.H.

Additional Information
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Jones-Odeh, E. *et al.* The correlation between cognitive performance and retinal nerve fibre layer thickness is largely explained by genetic factors. *Sci. Rep.* **6**, 34116; doi: 10.1038/srep34116 (2016).

AMERICAN ACADEMY™ OF OPHTHALMOLOGY

Correspondence

Re: Sitko et al.: Pitfalls in the use of stereoacuity in the diagnosis of nonorganic visual loss

(Ophthalmology 2016;123:198-202)

TO THE EDITORS: The study by Sitko et al[1] improves our understanding of the use of the Titmus Stereotest in helping to confirm nonorganic visual loss. However, in cases with possible medicolegal consequences, we believe that testing should be performed with the 3 lines of animals rather than the circles because the asymmetric circles provide monocular clues that can allow subjects to make a selection indicating stereopsis without stereopsis actually present, confounding the test.[2] The 3 rows of animals cannot discriminate higher levels of stereopsis, but testing for profound feigned visual loss with them reduces the chance of finding "false-positive" stereopsis, because they do not present any monocular clues.

ARGYRIOS CHRONOPOULOS, MD
JAMES SCOTT SCHUTZ, MD

Department of Ophthalmology, University Hospitals and School of Medicine, Geneva, Switzerland

Financial Disclosures: The authors have no proprietary or commercial interest in any materials discussed in this article.

Correspondence:
Argyrios Chronopoulos, MD, Department of Ophthalmology, University Clinics of Geneva and School of Medicine, Rue Alcide-Jentzer 22, 1205 Geneva, Switzerland. E-mail: argyrios.chronopoulos@hcuge.ch.

References

1. Sitko KR, Peragalo HJ, Bidot S, et al. Pitfalls in the use of stereoacuity in the diagnosis of nonorganic visual loss. Ophthalmology 2015;123:198–202.
2. Fricke TR, Siderov J. Stereopsis, stereotests, and their relation to vision screening and clinical practice. Clin Exp Optom 1997;80:165–72.

REPLY: We appreciate the recent critical reading of and response to our paper[1] by Chronopoulos and Schutz. They correctly note in their letter that the Wirt circles portion of the test provides the subject with monocular cues that can aid in making the correct choice, regardless of stereoacuity. We agree and mention in our paper that this is a limitation of the Titmus test, but chose it for our study given its wide use and prominent role in the diagnosis of nonorganic visual loss. However, Chronopoulos and Schutz also suggest that the animal figures section of the test does not suffer from the same limitations related to monocular cues that the circles section does, but this is not the case. The first 2 animals (400 and 200 arcseconds) have similar "double printing" monocular cues to those of the first through third circles and can be easily guessed monocularly. Furthermore, the maximal stereoacuity that can be determined from the animals is only 100 arcseconds. This level of stereoacuity corresponds to the fifth Wirt circle, which we found to correspond with a 95% probability of visual acuity of about 20/80, about one-half the acuity prediction of someone who can correctly identify 9/9 circles. We would encourage the use of alternative tests without such monocular cues (e.g., random dot stereotests) for the evaluation of nonorganic visual loss because they will likely help to reduce variability in the prediction of visual acuity. However, those tests should also be evaluated rigorously in a fashion similar to our evaluation of the Titmus test owing to other sources of variability.

KEVIN SITKO, MD[1]
JASON H. PERAGALLO, MD[2,3]
NANCY J. NEWMAN, MD[2,4,5]
BEAU B. BRUCE, MD, PHD[2,4,6]

[1]*Department of Surgery, University of New Mexico, Albuquerque, New Mexico;* [2]*Department of Ophthalmology, Emory University, Atlanta, Georgia;* [3]*Department of Pediatrics, Emory University, Atlanta, Georgia;* [4]*Department of Neurology, Emory University, Atlanta, Georgia;* [5]*Department of Neurological Surgery, Emory University, Atlanta, Georgia;* [6]*Department of Epidemiology, Emory University, Atlanta, Georgia*

Financial Disclosures: The authors have no proprietary or commercial interest in any materials discussed in this article.

Correspondence:
Kevin Sitko, MD, Department of Surgery, University of New Mexico, 2211 Lomas Boulevard NE, Albuquerque, NM 87106. E-mail: kevinsitko@gmail.com.

Reference

1. Sitko KR, Peragallo JH, Bidot S, et al. Pitfalls in the use of stereoacuity in the diagnosis of nonorganic visual loss. Ophthalmology 2016;123:198–202.

Re: Datiles et al.: Longitudinal study of age-related cataract using dynamic light scattering: loss of α-crystallin leads to nuclear cataract development

(Ophthalmology 2016;123:248-54)

TO THE EDITORS: We read with great interest the recent article by Datiles et al[1] exploring whether a longitudinal dynamic light scatter measurement of changes in unbound α-crystallin concentration within the lens predicts cataract progression. This study is built on a previous cross-sectional study for early detection of α-crystallin changes in lens using the same method.[2] We wondered whether brunescence and nuclear opalescence were graded separately, because they both are reflected within the term "nuclear cataract" and whether dynamic light scatter measurements reflect one more than the other, or are influenced by brunescence? Similarly, the published paper does not comment on smoking status, which is a major environmental influence on nuclear cataract, and we

Ophthalmology Volume 123, Number 8, August 2016

wondered whether the authors had any insights into whether smoking status differed in each tertile of the α-crystallin index, and might increase unbound α-crystallin, thereby causing cataract?

Given that common genetic variants in the *CRYAA* gene were recently associated with nuclear cataract formation in Asians,[3] we were wondering whether the authors have DNA or genetic data available for the participants in the study, to explore associations between depletion of α-crystallin and genetic variants in *CRYAA* gene. We have measured nuclear cataract in 324 female twins (TwinsUK cohort) at 2 time points (on average 9.7 years apart) by calculating the pixel density in the center of the nucleus from Scheimpflug images as previously described.[4] We have previously nominally replicated the association between *CRYAA* variants and cross-sectional nuclear cataract ($P = 0.01$ at rs870137).[5] We explored whether this same variant was also associated with cataract progression in the 324 twins. This was done using a score-based test (MERLIN), adjusting for age. We report here that we also find an association between rs870137 and cataract progression ($P = 0.01$). Our data suggest that common variants in the *CRYAA* gene may influence the prevalence and progression of nuclear cataract, and it would be fascinating to find out whether these variants influence the α-crystallin index.

Ekaterina Yonova-Doing, MSc[1]
Christopher J. Hammond, MD, PhD[1,2]

[1]*Department of Twin Research and Genetic Epidemiology;* [2]*Department of Ophthalmology, King's College London, London, UK*

Financial Disclosures: The authors have no proprietary or commercial interest in any materials discussed in this article.

Correspondence:
Christopher J. Hammond, MD, PhD, King's College London, Twin Research and Genetic Epidemiology, St. Thomas Hospital, Westminster Bridge Road, 3rd Floor, London, UK. E-mail: chris.hammond@kcl.ac.uk.

References

1. Datiles MB 3rd, Ansari RR, Yoshida J, et al. Longitudinal study of age-related cataract using dynamic light scattering: loss of alpha-crystallin leads to nuclear cataract development. Ophthalmology 2016;123:248–54.
2. Datiles MB 3rd, Ansari RR, Suh KI, et al. Clinical detection of precataractous lens protein changes using dynamic light scattering. Arch Ophthalmol 2008;126:1687–93.
3. Liao J, Su X, Chen P, et al. Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract. Hum Mol Genet 2014;23:6119–28.
4. Hammond CJ, Snieder H, Spector TD, Gilbert CE. Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. N Engl J Med 2000;342:1786–90.
5. Yonova EH, Nag A, Hysi PG, et al. Genome-wide association study of nuclear cataract finds suggestive association with a common variant in TRPM3. Invest Ophthalmol Vis Sci 2015;56:5815.

REPLY: We thank Drs Ekaterina Yonova-Doing and Christopher J. Hammond for their interest in our paper[1] and would like to respond to their questions.

With regard to lens brunescence, we used the Age-Related Eye Disease Study Cataract Clinical Grading System, which in contrast with the Lens Opacities Classification System, does not grade brunescence separately from opalescence. Brunescence is typically a later feature in nuclear cataract. The advantage of the dynamic light scattering system is that it can detect lens protein changes in the precataractous stages, when the lens is still transparent. Thus, at the time when dynamic light scattering is most useful, brunescence will generally not be an issue.

We agree that smoking is associated with increased risk of nuclear sclerosis and we would expect smoking to be associated with a decrease in α-crystallin index. However, we did not assess risk factors and did not know the smoking status of the individual patients. We did not perform genetic analysis in this study, but it could be an important feature in future studies.

Manuel B. Datiles, III, MD[1]
Frederick Ferris, III, MD[1]
Rafat R. Ansari, PhD[2]
J. Samuel Zigler, Jr., PhD[3]

[1]*Office of the Clinical Director, National Eye Institute, National Institutes of Health, Bethesda, Maryland;* [2]*National Aeronautics and Space Administration (NASA), Glenn Research Center, Cleveland, Ohio;* [3]*Wilmer Eye Institute, Johns Hopkins Hospital, Baltimore, Maryland*

Financial Disclosures: The authors have no proprietary or commercial interest in any materials discussed in this article.

Correspondence:
Manuel B. Datiles III, MD, Office of the Clinical Director, National Eye Institute, National Institutes of Health, 10 Center Drive, Building 10, Room 1, Bethesda, MD 20892-1860. E-mail: Datilesm@nei.nih.gov.

Reference

1. Datiles MB 3rd, Ansari RR, Yoshida J, et al. Longitudinal study of age-related cataract using dynamic light scattering: loss of α-crystallin leads to nuclear cataract development. Ophthalmology 2016;123:248–54.

Re: Jabbarvand et al.: Endophthalmitis occurring after cataract surgery: outcomes of more than 480 000 cataract surgeries, epidemiologic features, and risk factors

(Ophthalmology 2016;123:295-301)

TO THE EDITORS: This correspondence concerns the results of the study by Jabbarvand et al.[1] As per the results, 112 of 480 104 cases that underwent cataract surgery at a single hospital between 2006 and 2014 developed endophthalmitis. However, of the 25 920 cases where intracameral cefuroxime was injected, none developed endophthalmitis. This study also found that short-term pretreatment with topical or systemic antibiotics or postoperative subconjunctival injection of antibiotics reduced the odds of endophthalmitis by 40% to 50%, but the difference was not significant ($P = 0.2$). The authors concluded from their results that intracameral cefuroxime was 100% effective in preventing postcataract surgery endophthalmitis. We would like to put forth our observations regarding 2 statistical concepts, sample size, and risk reduction.

AMERICAN ACADEMY™ OF OPHTHALMOLOGY

Genetic and Dietary Factors Influencing the Progression of Nuclear Cataract

Ekaterina Yonova-Doing, MSc,[1] *Zoe A. Forkin, BSc,*[1,3] *Pirro G. Hysi, MD, PhD,*[1] *Katie M. Williams, MPhil, FRCOphth,*[2] *Tim D. Spector, MD, PhD,*[1] *Clare E. Gilbert, FRCOphth, MD,*[4] *Christopher J. Hammond, MD, FRCOphth*[1,2]

Purpose: To determine the heritability of nuclear cataract progression and to explore prospectively the effect of dietary micronutrients on the progression of nuclear cataract.

Design: Prospective cohort study.

Participants: Cross-sectional nuclear cataract and dietary measurements were available for 2054 white female twins from the TwinsUK cohort. Follow-up cataract measurements were available for 324 of the twins (151 monozygotic and 173 dizygotic twins).

Methods: Nuclear cataract was measured using a quantitative measure of nuclear density obtained from digital Scheimpflug images. Dietary data were available from EPIC food frequency questionnaires. Heritability was modeled using maximum likelihood structural equation twin modeling. Association between nuclear cataract change and micronutrients was investigated using linear and multinomial regression analysis. The mean interval between baseline and follow-up examination was 9.4 years.

Main Outcome Measures: Nuclear cataract progression.

Results: The best-fitting model estimated that the heritability of nuclear cataract progression was 35% (95% confidence interval [CI], 13−54), and individual environmental factors explained the remaining 65% (95% CI, 46−87) of variance. Dietary vitamin C was protective against both nuclear cataract at baseline and nuclear cataract progression ($\beta = -0.0002$, $P = 0.01$ and $\beta = -0.001$, $P = 0.03$, respectively), whereas manganese and intake of micronutrient supplements were protective against nuclear cataract at baseline only ($\beta = -0.009$, $P = 0.03$ and $\beta = -0.03$, $P = 0.01$, respectively).

Conclusions: Genetic factors explained 35% of the variation in progression of nuclear cataract over a 10-year period. Environmental factors accounted for the remaining variance, and in particular, dietary vitamin C protected against cataract progression assessed approximately 10 years after baseline. *Ophthalmology 2016;123:1237-1244 © 2016 by the American Academy of Ophthalmology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).*

Age-related cataract is the leading cause of blindness in the world, affecting approximately 20 million people, particularly in sub-Saharan Africa.[1] Its prevalence increases from 2.9% in the 43- to 54-year age group to 40% in those older than 75 years of age.[2] As the world's population ages, cataract will remain a serious healthcare and socioeconomic burden, in terms of both healthcare provision and blindness in less-developed countries.

Nuclear cataract is the most common form of age-related cataract.[2] Apart from age, other factors associated with nuclear cataract are smoking, oxidative stress, and dietary antioxidant intake.[3−5] However, studies of the effect of dietary vitamin C intake,[6−11] serum vitamin C concentrations,[6,9,11−13] and vitamin C supplementation[6,10,14] on nuclear cataract formation often have given conflicting results. Case-control studies[7,11,12,14] and some cohort studies[6,9,10] have found protective effects. Other prospective cohort studies have found no effect overall[8,13,15] or protective effects only in subgroups.[8,15] Similarly to vitamin C, dietary[6,16] and supplemental[14,17] vitamin E intake and vitamin E blood concentrations[6,13] have been shown to be related inversely with nuclear cataract. Randomized clinical trials of vitamins C and E supplementation alone or in combination with other vitamins failed to find an effect.[18,19] Vitamin A has been associated with a reduced risk of nuclear cataract,[9,20,21] as have lutein and zeaxanthin.[22−24] The studies exploring dietary nutrients and cataract progression have findings similar to those looking at prevalent cataract, with cohort studies finding a protective effect.[16,25] However, supplement trials largely have failed to find an effect.[18,26,27]

As opposed to vitamins and micronutrients,[28] the role of minerals in cataract formation in general and in nuclear cataract in particular is poorly studied. Together with epidemiologic factors, genetic factors also play a role in cataract formation. We previously reported that genetic factors explain 48% of cross-sectional variance in age-related nuclear cataract.[29] In a recent genome-wide meta-analysis, variants in 2 genes, *CRYAA* and *KCNAB1*, were found to be associated with nuclear cataract in Asian populations,[30] but no findings are available for populations of European origin. In comparison with epidemiologic factors, little is known about genetic susceptibility factors in age-related cataract.

http://dx.doi.org/10.1016/j.ophtha.2016.01.036
ISSN 0161-6420/16

Ophthalmology Volume 123, Number 6, June 2016

Factors that lead to the development of a phenotype may be different from factors underlying change, such as progression of lens opacity. Therefore, we set out to establish the relative importance of genes on progression of nuclear cataract using a classic twin model with a highly quantitative measure of nuclear cataract. We also examined how intake of micronutrients and supplements associated with nuclear cataract at baseline affects nuclear cataract progression over a decade.

Methods

Subjects

Nuclear cataract data at baseline were available for 2515 white female twins (mean age, 62.3 years; range, 50.1−83.1 years) from the TwinsUK cohort, 2054 of whom had also completed a food frequency questionnaire (FFQ) around the time of their eye examination. The median time interval between an eye test taking place and a FFQ completion was 2 years. The 461 twins with cataract data but without FFQ data were 2.5 years younger on average and less affected by cataract. Cataract progression data were collected from 324 twins (151 monozygotic [MZ] and 173 dizygotic [DZ]) with a mean age at follow-up of 69.8±5.4 years (range, 58.3−83.6 years) as part of the Healthy Ageing in Twins (HATS) study between 2006 and 2010.[31] Individuals included in the follow-up were all part of our original cataract heritability study of 1012 twin participants assessed in 1998 and 1999.[29] The mean time between baseline and second visits was 9.4 years (range, 7−12 years). The smaller number of individuals with follow-up data is mainly due to the fact that the HATS study (in which the follow-up data were collected) was not designed specifically as a cataract follow-up study and had different selection criteria: Participants were aged more than 40 years and had to have previously attended clinical phenotyping irrespective of whether they had an eye examination or not (N = 4610) (Fig 1). The TwinsUK study started in 1992, but eye measures were performed only on subjects more than 50 years of age in 1998−1999, and subsequently from 2006. That meant that individuals (aged ≥50 years) who attended the HATS visit and who did not have eye examinations in 1989−1999 had their baseline cataract assessment during HATS (2006−2011; N = 1523). Reasons for having only longitudinal data for 324 of the original 1012 twins included death (N = 52), withdrawal of participation from the TwinsUK registry (N = 169), noncontactable (N = 30), refusal of further phenotyping (N = 82), cataract surgery (N = 11), and refusal of dilating drops or unavailability of ophthalmic testing at the HATS visit (N = 344).

Both the baseline study and the HATS study received local research ethics approval and were conducted according to the tenets of the Declaration of Helsinki. All the participants gave written informed consent.

Phenotyping

Nuclear Cataract Scores. Digital black and white lens photographs were taken using a Scheimpflug camera (Case 2000; Marcher Enterprises Ltd., Worcester, UK), and the same camera was used at both baseline and follow-up. Nuclear cataract was measured quantitatively by calculating the pixel density in the center of the lens nucleus, also known as the central nuclear dip score (NDS).[29] This score measures the amount of white scatter (opalescence), and more opacification results in higher pixel density. Because NDS uses black-and-white images, it does not assess the brunescence of the lens. Nuclear cataract progression was measured as the difference in measurements between the visits: ΔNDS = NDS at follow-up − NDS at baseline. Both NDS and ΔNDS were not normally distributed and therefore were transformed using natural logarithm before the analysis.

Nutrient Intake. Intake of micronutrients (vitamins and minerals) and supplements was estimated using the EPIC FFQ, which was self-administered at the baseline visit. This questionnaire explored the average frequency of intake of 131 foods and supplements over a 1-year period.[32,33] Nutrient intake was calculated using an established nutrient database, and the dietary variables were adjusted for calorie intake, yielding an energy-adjusted mg/μg of each nutrient per person per day.[32,34,35] We considered the following micronutrients in the analysis: sodium, potassium, calcium, magnesium, phosphorus, iron,

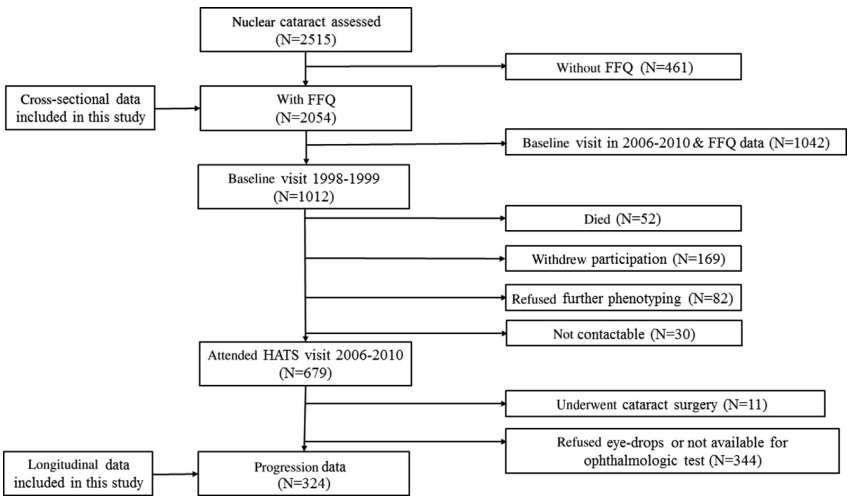


Figure 1. Consort diagram of the study showing the number of individuals who participated in the different parts of the study and reasons for no participation at follow-up. FFQ = food frequency questionnaire; HATS = Healthy Ageing in Twins.

copper, zinc, chloride, manganese, iodine, retinol, carotene, vitamin D, vitamin E, thiamine, riboflavin, niacin, tryptophan, vitamin B6, vitamin B12, folate, pantothenate, biotin, and vitamin C.

Data on supplement intake were available for 33 different supplements. However, the percentage of individuals taking any single supplement was 10% or less. Supplements were grouped as follows: *any supplements*; *micronutrient supplements* (vitamins and minerals in any combination); *micronutrient supplements excluding multivitamins* (e.g., vitamin C only, vitamin D only, iron only, ACD complex); *minerals only* (e.g., iron only, calcium only); and *other supplements* (e.g., aloe vera, Echinacea, Ginkgo, omega-3). Each supplement group was coded as a binary variable, with "yes" indicating that they took 1 or more of the supplements in a specific group.

Statistical Analysis

Modeling of Heritability. Heritability was analyzed in 310 twins (155 pairs: 72 MZ and 83 DZ) because data were missing on 14 cotwins. Zygosity was determined by a standardized questionnaire and confirmed using genome-wide single nucleotide polymorphism genotyping data or DNA short tandem repeat fingerprinting.

Twin studies are able to estimate the heritability of a trait (the amount of variance explained by genetic factors) using maximum likelihood structural equation modeling. The variance of the trait and the covariance within twin pairs are used to estimate additive genetic effects (A), shared/family environmental effects (C), and individual environmental effects (E). We implemented the modeling in the OpenMx package (http://openmx.psyc.virginia.edu). The goodness of fit of the full ACE model and submodels were compared with the observed data and the best fitting model was selected.

Nutrient Factor Analysis. Comparisons of means and proportions for all variables between individuals with or without follow-up data, or between MZ and DZ twins per group in terms of age, nuclear cataract scores, and nutrient and supplement intake were performed using 2-sample, 2-tailed *t* tests or *z* tests, assuming equal variance.

Association was assessed using linear regression analyses. Univariable linear regression was first carried out where each factor or supplement group was individually regressed against NDS at baseline. All nutrients or supplement groups showing significant univariable association ($P < 0.05$) were then included in a multivariable linear regression model; independent variables were identified using a stepwise backward procedure with threshold for removal set at 0.05. Factors showing significant ($P < 0.05$) association in the multivariable model were tested for association with progression. We used linear models to establish the relationship between NDS (continuous variable) and nutrients, but because NDS had to be normalized, giving a clinical interpretation of the betas becomes more difficult. Therefore, in addition to the linear models, we calculated risk reduction by calculating relative risk ratios (RRRs) using multinomial regression. In this case, NDS, ΔNDS, and the associated nutrients were divided into tertiles, and the first tertile was set as reference while supplement intake per supplement group was kept binary. In all cases, models were adjusted for family structure and age at the first visit only (baseline analysis) or for both age at baseline and Δage (age at follow-up − age at baseline). All analyses were carried using the STATA10 statistical package (StataCorp LP, College Station, TX; www.stata.com).

Results

Cross-sectional data were available for 2054 white female twins (827 MZ and 916 DZ), 324 (151 MZ and 173 DZ) of whom also had nuclear cataract measured at follow-up. Baseline characteristics and nutrient and supplement intake are shown in Table 1, and an example of a lens image is available in Figure 2. The twins with follow-up data were on average 1.1 years younger at baseline (60.4 vs. 61.5 years) and, given their younger age, had less cataract (mean NDS scores of 55.3 and 60.4, respectively) compared with those with only cross-sectional data. In both cases, these differences were not statistically significant ($P > 0.05$). The MZ and DZ twins with follow-up data were similar in terms of age and NDS scores ($P > 0.05$). The MZ and DZ twins with cross-sectional data only were similar in terms of age, but the MZ twins had a slightly higher NDS score (61.6 vs. 59.3; $P = 0.02$).

There were also no statistically significant differences between groups in terms of micronutrient intake except for iron ($P = 0.02$), thiamine ($P = 0.04$), and biotin ($P = 0.01$). The twins with follow-up data had slightly lower iron and thiamine intake (mean of 12.6 and 1.7 mg, respectively) and slightly higher biotin intake (mean of 49.7 mg) compared with individuals without follow-up data. There were also no significant differences in supplement intake between the 2 groups ($P > 0.05$). There were no statistically significant differences between MZ and DZ twins in terms of nutrient or supplement intake ($P > 0.05$).

As expected, nuclear cataract scores progressed in all participants (Fig 3). The mean baseline central NDS was 55±11 (range, 32−99) with the score increasing by an average of 19.9±16.9 (range, 1−137) over the period of follow-up. The heritability analysis, conducted on 155 twin pairs (72 MZ and 83 DZ pairs), showed that the best-fitting model was one explained by additive genetic factors and the unique (individual) environment, with no significant effect of a common environment or nonadditive genetic factors. Calculations estimated the heritability to be 0.35, meaning that genetic factors explained 35% of variance (95% confidence interval [CI], 13−54) in progression of nuclear cataract, with individual environmental factors accounting for the remaining 65% (95% CI, 46−87).

To test associations between micronutrient intake and cataract progression, we used univariable regression (Table 2) followed by stepwise regression in 2054 female twins who had baseline data on nutrient intake. Seven micronutrients showed a significant association ($P < 0.05$) with NDS and were used in multivariable analysis: potassium, magnesium, manganese, phosphorus, vitamins C and E, and folate. After stepwise multivariable regression, 2 factors remained significantly associated with NDS at baseline: vitamin C ($\beta = -0.0002$; standard deviation [SD] = 6.3E-05; $P = 0.01$) and manganese ($\beta = -0.009$; SD = 0.04; $P = 0.03$). From these 2 nutrients, only vitamin C showed association with cataract progression ($\beta = -0.001$; SD = 0.001; $P = 0.03$). A sensitivity analysis excluding subjects with greatest progression (>100 units of change) did not alter the result. Comparing people in the highest and lowest tertiles of vitamin C intake was associated with a 19% risk reduction at baseline (RRR, 0.81; 95% CI, 0.68−0.96) and a 33% risk reduction of cataract progression (RRR, 0.66; 95% CI, 0.47−0.91) (Table 3). Manganese intake was associated with 20% risk reduction (RRR, 0.80; 95% CI, 0.67−0.95) at baseline (Table 3).

Two supplement groups, *micronutrient supplements* and *minerals only*, showed a significant association with NDS ($P < 0.05$) (Table 2), but only *micronutrient supplements* stayed significant in the multivariate model ($\beta = -0.03$; SD = 0.01; $P = 0.01$), and their intake led to an 18% risk reduction in people within the highest compared with the lowest tertile of nutrient intake (RRR, 0.82; 95% CI, 0.57−1.20) (Table 3). We found no statistically significant association between taking micronutrients in supplemental form and progression of nuclear cataract.

Ophthalmology Volume 123, Number 6, June 2016

Table 1. Baseline Sample Characteristics and Nutrient Intakes in Individuals with or without Follow-up Data

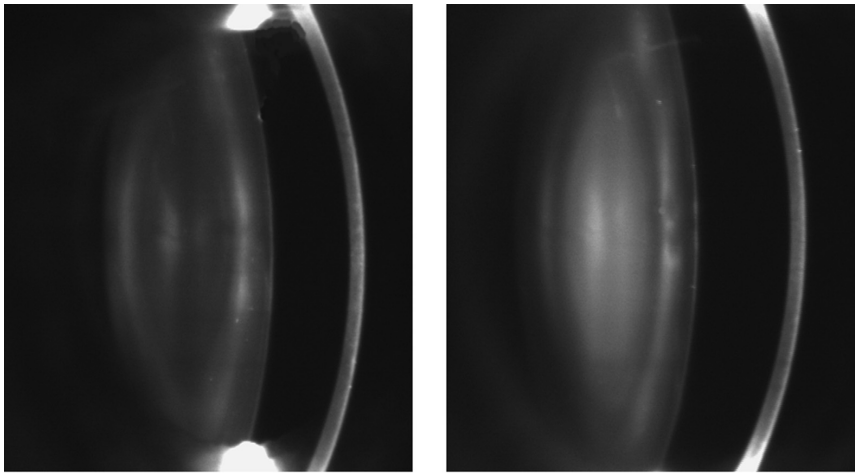
	Subjects without Follow-up			Subjects with Follow-up		
	Total	*MZ*	*DZ*	*Total*	*MZ*	*DZ*
No. of individuals	1730	827	916	324	151	173
Zygosity ratio (MZ:DZ)	1:1.1	-	-	1:1.2	-	-
Age (mean ± SD)	61.5±6.5	61.7±6.7	61.4±6.4	60.4±5.1	60.8±5.5	60.0±5.2
NDS (mean ± SD)	60.4±17.2	61.3±17.4	59.0±14.2	55.3±11.2	55.3±11.4	55.3±11.1
Sodium (mg)	2262.8±508.7	2265.3±476.3	2258.7±535.6	2237.4±456.4	2227.7±444.4	2247.2±444.4
Potassium (mg)	4013.5±637.4	3997.0±622.4	4026.9±650.6	4033.7±580.5	4094.5±588.4	3972.5±469.4
Calcium (mg)	1117.1±284.7	1118.5±284.9	1125.1±284.6	1118.9±291.5	1138.3±295.0	1099.4±568.0
Magnesium (mg)	347.3±56.4	347.3±56.8	347.2±56.0	343.8±55.0	347.0±58.0	340.6±287.5
Phosphorus (mg)	1527.1±247.0	1527.1±234.9	1527.1±257.8	1522.0±239.3	1532.0±251.0	1512.1±227.5
Iron (mg)*	13.1±3.0	13.2±3.2	13.0±2.8	12.6±2.6	12.5±2.7	12.7±2.5
Copper (mg)	1.5±0.5	1.5±0.6	1.5±0.4	1.5±0.4	1.5±0.4	1.6±0.5
Zinc (mg)	10.2±1.7	10.2±1.8	10.1±1.7	10.2±1.7	10.2±1.8	10.1±1.6
Chloride (mg)	3629.6±792.9	3633.6±749.4	3623.0±828.6	3578.0±721.3	3566.7±690.3	3589.4±753.3
Manganese (mg)	4.2±1.2	4.1±1.1	4.2±1.2	4.2±1.1	4.3±1.1	4.2±1.1
Iodine (mg)	225.0±75.8	224.2±75.2	225.8±76.5	229.2±64.2	230.0±61.4	228.5±67.2
Retinol (μg)	579.5±817.8	569.1±570.6	554.8±496.6	611.8±472.9	588.2±422.6	635.6±519.0
Carotene (μg)	5343.4±3067.4	5503.7±3263.8	5200.4±2874.9	5305.6±3915.4	5663.8±4823.8	4945.0±2679.4
Vitamin D (μg)	2.7±1.4	2.7±1.1	2.6±1.5	2.8±1.1	3.0±1.0	2.6±1.0
Vitamin E (mg)	11.5±3.2	11.6±3.4	11.4±3.1	11.7±3.4	11.9±3.6	11.5±3.2
Thiamine (mg)*	1.8±0.4	1.8±0.4	1.8±0.4	1.7±0.3	1.7±0.3	1.7±0.3
Riboflavin (mg)	2.5±0.7	2.4±0.7	2.5±0.7	2.4±0.6	2.5±0.6	2.4±0.7
Niacin (mg)	22.0±5.7	22.2±5.1	21.8±6.2	21.3±4.5	21.3±4.6	21.2±4.4
Tryptophan (mg)	17.4±3.0	17.5±2.7	17.3±3.3	17.2±2.5	17.3±2.5	17.1±2.6
Vitamin B6 (mg)	2.6±0.6	2.6±0.6	2.5±0.5	2.5±0.5	2.5±0.5	2.5±0.5
Vitamin B12 (μg)	6.5±3.2	6.7±3.6	6.4±2.9	6.7±2.3	6.7±2.3	6.7±2.4
Folate (μg)	402.2±113.1	400.7±114.0	403.2±112.3	395.7±98.9	402.0±95.9	389.4±101.8
Pantothenate (mg)	7.4±16.0	7.5±21.3	7.2±8.6	6.8±4.2	6.5±2.1	7.1±5.6
Biotin (mg)*	48.1±10.5	47.7±10.3	48.5±10.8	49.7±10.3	50.6±10.2	48.7±10.3
Vitamin C (mg)	165.1±73.9	167.6±74.2	163.0±73.7	166.8±65.0	166.9±68.1	166.7±65.0
Any supplement (%)	55.1	54.8	55.4	55.0	54.1	55.9
Micronutrients (%)	32.57	32.4	33.2	31.7	32.8	30.8
Micronutrients excluding multivitamins (%)	23.6	24.1	23.2	21.6	24.2	19.3
Minerals only (%)	7.4	7.8	7.0	6.9	6.4	7.2
Other supplements (%)	44.9	46.2	44.4	47.1	44.2	49.5

DZ = dizygotic; MZ = monozygotic; NDS = nuclear dip score; SD = standard deviation.
The baseline characteristics of the participants and the baseline intake of micronutrients (mean ± standard deviation [SD]) and supplements per supplement group (% of users) are shown. The supplement groups studied are as follows: any supplement, micronutrient supplements (vitamins and mineral in any combination), micronutrient supplements excluding multivitamins (e.g., vitamin C only, vitamin D only, iron only, ACD complex), minerals only (e.g., iron only, calcium only), and other supplements (e.g., aloe vera, Echinacea, Ginkgo, omega-3).
*Denotes statistically significant difference ($P < 0.05$) between subjects with and without and without follow-up.

Discussion

This study found that progression of nuclear cataract over a 10-year period in a group of UK female twins is influenced by genetic factors that explain 35% of variance. The heritability estimate of cataract progression is lower than our previous cross-sectional estimates of susceptibility to development of nuclear cataract in this cohort,[29] and it is also lower than the heritability estimated in the 324 individuals estimated from the nuclear score measurement at follow-up (61%; 95% CI, 45–72). This is consistent with previous studies showing that heritability generally is lower when examining change, compared with cross-sectional studies.[36–38] In addition to early developmental differences and the body's response to environmental factors in adulthood, environmentally driven processes or accumulated "errors" (e.g., somatic gene mutation and epigenetic remodeling) might play a greater role in determining change during aging than genetic factors.[38]

This study also identified vitamin C as a micronutrient affecting nuclear cataract progression. We also replicate the previously found association between cross-sectional cataract and vitamin C intake. Vitamin C intake has long been studied in relation to age-related cataract because it is the L-enantiomer of ascorbate. A significant concentration of ascorbate is present in the aqueous humor that bathes the lens and may reduce oxidation products in the lens, thus reducing oxidative stress.[39,40] However, the conclusions of the many studies of the effects of ascorbate on cataract development are inconsistent and often conflicting.[6–15] Many of these have studied relatively well-nourished populations and are cross-sectional, although cross-sectional studies in India, where overall antioxidant levels may be lower, have found an inverse relationship between vitamin



Relatively clear nucleus | Nuclear cataract

Figure 2. Black-and-white Scheimpflug lens images of a healthy lens (*left*) and a lens with nuclear cataract (*right*). The center of the lens (lens nucleus) on the *right* is much whiter than the one on the *left*.

C and cataract.[9,20] Our results are similar to those of the Carotenoids in Age-Related Eye Disease Study that showed vitamin C intake, assessed with an FFQ 10 years before cataract assessment, to be protective of nuclear cataract prevalence.[15] The Blue Mountains Eye Study also found that vitamin C intake, through both diet and supplements together, resulted in a lower nuclear cataract incidence over 10 years.[10] This study is the first, to our knowledge, to show that dietary vitamin C intake protects against progression of nuclear lens opacity.

We also found dietary manganese to be protective against cross-sectional nuclear cataract independent of vitamin C.

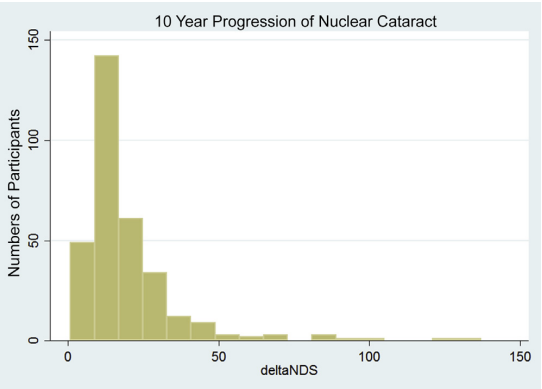


Figure 3. Progression of nuclear cataract between the 2 visit dates. Graphical representation of the progression of nuclear cataract (deltaNDS) between the 2 visits. The y-axes show frequency of deltaNDS per bins with width of 6.25 points. deltaNDS = NDS at follow-up − NDS at baseline; NDS = nuclear dip score.

We cannot exclude that this association was a type I error, given we did not find an association between dietary manganese and nuclear cataract progression and the lack of a dose response (Table 3), although factors associated with incidence and progression do not always overlap. Manganese is an important antioxidant present in the human lens,[41–43] and its concentration has been reported to be lower in cataractous lenses in comparison with normal lenses.[43,44] This study was not designed to elucidate the cause–effect relationship underlying the associations we found; therefore, we cannot distinguish whether manganese depletion is a cause or effect of cataractogenesis. Further studies are needed to answer this question. We also detected an association between supplemental intake of micronutrients and cross-sectional nuclear cataract but not between supplemental nutrients and cataract progression. These results are similar to those reported in the Blue Mountains Eye Study.[45] Because only 10% or less of the participants in our study took any single supplement, we had to group supplements together; therefore, we could not draw conclusions on the effect of any single supplement or of components of supplements (e.g., supplemental vitamin C).

We used a highly quantitative measure of cataract from digital images (NDS), which essentially measures the nuclear opalescence (or "white scatter") of the lens. The measure also was highly reproducible: the intraclass correlation coefficient for the worse eye in 30 subjects from our original study[29] who came for repeat measurements was 0.93. The fact that every subject measured showed progression suggests that NDS is sensitive to change. Many epidemiologic studies have used the Lens Opacity Classification System (LOCS) grading scale, comparing phenotype with standardized photographs of 6 stages of lens opacification, which includes both nuclear opalescence and nuclear color or brunescence.[4] The LOCS III

Ophthalmology Volume 123, Number 6, June 2016

Table 2. Results from Univariable Regression Models

	Beta	Standard Error	P Value
	Micronutrients		
Sodium (mg)	5.41E-06	9.58E-06	0.56
Potassium (mg)*	−1.58E-05	7.54E-06	0.04
Calcium (mg)	−1.95E-05	1.52E-05	0.20
Magnesium (mg)*	−0.010	0.004	0.01
Phosphorus (mg)*	−4.01E-05	1.94E-05	0.04
Iron (mg)	−1.15E-04	0.002	0.95
Copper (mg)	0.001	0.008	0.86
Zinc (mg)	−7.76E-04	0.003	0.77
Chloride (mg)	3.79E-06	6.10E-06	0.53
Manganese (mg)*	−0.010	0.004	0.01
Iodine (mg)	−1.10E-04	6.07E-05	0.07
Retinol (μg)	2.36E-06	3.90E-06	0.55
Carotene (μg)	−1.67E-06	1.40E-06	0.23
Vitamin D (μg)	−0.004	0.003	0.22
Vitamin E (mg)*	−0.003	0.001	0.04
Thiamine (mg)	−0.013	0.013	0.30
Riboflavin (mg)	−0.011	0.006	0.08
Niacin (mg)	−1.10E-04	8.26E-04	0.89
Tryptophan (mg)	−0.001	0.001	0.27
Vitamin B6 (mg)	−0.002	0.009	0.81
Vitamin B12 (μg)	−0.001	0.001	0.50
Folate (μg)*	−9.91E-05	4.06E-05	0.02
Pantothenate (mg)	−2.81E-05	1.87E-04	0.88
Biotin (mg)	−3.01E-04	4.17E-04	0.47
Vitamin C (mg)*	−1.742E-04	6.19E-05	0.01
	Supplement Groups†		
Any supplement	−0.015	0.009	0.12
Micronutrients*	−0.032	0.013	0.01
Micronutrients excluding multivitamins	−0.023	0.012	0.06
Minerals only*	−0.038	0.016	0.02
Any other supplement	0.005	0.014	0.72

The results of the univariable linear regression analysis between nuclear cataract (natural logarithm-transformed NDS) and energy-adjusted micronutrient intakes and between nuclear cataract and supplement intake per supplement group are shown.
*Denotes statistically significant associations at $P < 0.05$.
†Denotes that in the case of supplement groups, supplement intake was coded binary (presence vs. absence of intake of at least 1 of the components in the group). All analyses were adjusted for age and family structure.

Table 3. Results of Multinomial Regression Analysis for Factors Associated with Cross-Sectional Nuclear Cataract and Nuclear Cataract Progression

	Cross-Sectional Results			
	Vitamin C	*RRR*	*95% CI*	*P Value*
NDS Tertiles	34.5–53.2	reference		
	53.3–54.5	0.89	0.77–1.02	0.09
	54.6–229.2	0.81	0.68–0.96	0.01
	Manganese	*RRR*	*95% CI*	*P Value*
NDS Tertiles	34.5–53.2	reference		
	53.3–54.5	0.76	0.66–0.87	0.001
	54.6–229.2	0.8	0.67–0.95	0.01
	Micronutrients	*RRR*	*95% CI*	*P Value*
NDS Tertiles	34.5–53.2	reference		
	53.3–54.5	0.82	0.60–1.12	0.82
	54.6–229.2	0.82	0.57–1.20	0.82
	Progression Results			
	Vitamin C	*RRR*	*95% CI*	*P Value*
ΔNDS Tertiles	1.0–12.6	reference		
	12.7–19.3	0.75	0.54–1.04	0.09
	19.4–137.1	0.66	0.47–0.91	0.01

CI = confidence interval; NDS = nuclear dip score; ΔNDS = NDS at follow-up − NDS at baseline; RRR = relative risk ratio.
The results from the multinomial regression analysis for factors associated with cross-sectional (vitamin C and manganese) and progression (vitamin C) are shown. The RRR with its 95% CIs for each tertile of NDS or progression (ΔNDS) is reported. The minimum and maximum NDS score per tertile are reported.

was developed to increase steps between scores to allow greater sensitivity to change, accepting a lower intergrader reproducibility.[46] Longitudinal studies using the LOCS III scale show relatively little change: In the Longitudinal Study of Cataract, only 24% of participants had an increase in nuclear opacities over an average of 4.6 years.[25] Although our central NDS is not the same measure, it is highly correlated with average nuclear opalescence graded digitally or using a slit lamp.[29] Digital image–derived NDSs using pixel density counts may be better suited for measuring progression and allowed our study the power to detect associations with a relatively small sample size.

Study Limitations

A potential limitation is that our cohort is based on twin volunteers rather than a population study, but they are unselected and from across the United Kingdom and are unlikely to significantly differ from the UK general population.[47] Twin studies use the "equal environment assumption" that the degree of shared family environment is the same for both MZ and DZ twin pairs. This is generally found to be true, although there are few studies of elderly subjects that explore this assumption. In addition, the TwinsUK cohort is predominantly a female cohort, and we could not assess any gender differences in risk factors. The findings of this study can be generalizable only to Caucasian women of similar age because they reflect cataract progression in a group of white British women between, on average, the ages of 60 and 70 years, and so may not reflect other population groups or age ranges. In this article, we explored the effect of all micronutrients on nuclear cataract formation; however, we had no data on carotenoid (lutein and zeaxanthin) intake. We also lacked power to explore the effects of smoking on cataract progression because 85% of participants have never smoked.

Those participants with follow-up data collected were seen as part of the HATS study, which was not designed as a cataract follow-up study. This meant that the number of subjects decreased to 324 individuals, thus reducing the amount of data we could analyze and our power. The individuals who were lost to follow-up in HATS were in general of lower socioeconomic status, had higher self-rated health status, and were less health aware.[31] Any introduced bias probably would have resulted in loss of power because this group of individuals are more likely to have less healthy diets and more cataract. For this reason, we decided to test the association with progression only for nutrients that were associated with NDS at baseline. Those with follow-up data were on average 1.8 years younger than the

original cohort, but in general they were not significantly different in other respects or in nutrient or supplement intake, hopefully reducing potential selection bias in the progression data. As in any observational study, ours is potentially susceptible to residual confounding, missing data, or misspecification of variables.

In summary, this study has shown that progression of nuclear cataract over a 10-year period is influenced by genetic factors with a heritability of 35%. Dietary vitamin C and manganese, both factors related to oxidative stress, seem to influence cross-sectional nuclear cataract, and vitamin C intake also significantly influences nuclear cataract progression.

Acknowledgments. The authors thank all the twin volunteers for their participation and Alex MacGregor and Sue Fairweather-Tait at the University of East Anglia for their work in analysis of the FFQ data used in this study.

References

1. Pascolini D, Mariotti SPM. Global estimates of visual impairment: 2010. Br J Ophthalmol 2012;96:614–8.
2. Klein BE, Klein R, Lee KE. Incidence of age-related cataract: the Beaver Dam Eye Study. Arch Ophthalmol 1998;116:219–25.
3. Kelly SP, Thornton J, Edwards R, et al. Smoking and cataract: review of causal association. J Cataract Refract Surg 2005;31:2395–404.
4. Truscott RJ. Age-related nuclear cataract-oxidation is the key. Exp Eye Res 2005;80:709–25.
5. Chiu CJ, Taylor A. Nutritional antioxidants and age-related cataract and maculopathy. Exp Eye Res 2007;84:229–45.
6. Jacques PF, Chylack LT Jr, Hankinson SE, et al. Long-term nutrient intake and early age-related nuclear lens opacities. Arch Ophthalmol 2001;119:1009–19.
7. Leske MC, Chylack LT Jr, Wu SY. The Lens Opacities Case-Control Study. Risk factors for cataract. Arch Ophthalmol 1991;109:244–51.
8. Lyle BJ, Mares-Perlman JA, Klein BE, et al. Antioxidant intake and risk of incident age-related nuclear cataracts in the Beaver Dam Eye Study. Am J Epidemiol 1999;149:801–9.
9. Ravindran RD, Vashist P, Gupta SK, et al. Inverse association of vitamin C with cataract in older people in India. Ophthalmology 2011;118:1958–1965 e2.
10. Tan AG, Mitchell P, Flood VM, et al. Antioxidant nutrient intake and the long-term incidence of age-related cataract: the Blue Mountains Eye Study. Am J Clin Nutr 2008;87:1899–905.
11. Valero MP, Fletcher AE, De Stavola BL, et al. Vitamin C is associated with reduced risk of cataract in a Mediterranean population. J Nutr 2002;132:1299–306.
12. Ferrigno L, Aldigeri R, Rosmini F, et al. Associations between plasma levels of vitamins and cataract in the Italian-American Clinical Trial of Nutritional Supplements and Age-Related Cataract (CTNS): CTNS Report #2. Ophthalmic Epidemiol 2005;12:71–80.
13. Vitale S, West S, Hallfrisch J, et al. Plasma antioxidants and risk of cortical and nuclear cataract. Epidemiology 1993;4:195–203.
14. Robertson JM, Donner AP, Trevithick JR. A possible role for vitamins C and E in cataract prevention. Am J Clin Nutr 1991;53(1 Suppl):346S–51S.
15. Mares JA, Voland R, Adler R, et al. Healthy diets and the subsequent prevalence of nuclear cataract in women. Arch Ophthalmol 2010;128:738–49.
16. Jacques PF, Taylor A, Moeller S, et al. Long-term nutrient intake and 5-year change in nuclear lens opacities. Arch Ophthalmol 2005;123:517–26.
17. Chasan-Taber L, Willett WC, Seddon JM, et al. A prospective study of vitamin supplement intake and cataract extraction among U.S. women. Epidemiology 1999;10:679–84.
18. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss: AREDS report no. 9. Arch Ophthalmol 2001;119:1439–52.
19. Christen WG, Glynn RJ, Sesso HD, et al. Age-related cataract in a randomized trial of vitamins E and C in men. Arch Ophthalmol 2010;128:1397–405.
20. Dherani M, Murthy GV, Gupta SK, et al. Blood levels of vitamin C, carotenoids and retinol are inversely associated with cataract in a North Indian population. Invest Ophthalmol Vis Sci 2008;49:3328–35.
21. Cumming RG, Mitchell P, Smith W. Diet and cataract: the Blue Mountains Eye Study. Ophthalmology 2000;107:450–6.
22. Delcourt C, Carriere I, Delage M, et al. Plasma lutein and zeaxanthin and other carotenoids as modifiable risk factors for age-related maculopathy and cataract: the POLA Study. Invest Ophthalmol Vis Sci 2006;47:2329–35.
23. Karppi J, Laukkanen JA, Kurl S. Plasma lutein and zeaxanthin and the risk of age-related nuclear cataract among the elderly Finnish population. Br J Nutr 2012;108:148–54.
24. Liu XH, Yu RB, Liu R, et al. Association between lutein and zeaxanthin status and the risk of cataract: a meta-analysis. Nutrients 2014;6:452–65.
25. Leske MC, Chylack LT Jr, He Q, et al. Antioxidant vitamins and nuclear opacities: the longitudinal study of cataract. Ophthalmology 1998;105:831–6.
26. Gritz DC, Srinivasan M, Smith SD, et al. The Antioxidants in Prevention of Cataracts Study: effects of antioxidant supplements on cataract progression in South India. Br J Ophthalmol 2006;90:847–51.
27. Mathew MC, Ervin AM, Tao J, Davis RM. Antioxidant vitamin supplementation for preventing and slowing the progression of age-related cataract. Cochrane Database Syst Rev 2012;6:CD004567.
28. Weikel KA, Garber C, Baburins A, Taylor A. Nutritional modulation of cataract. Nutr Rev 2014;72:30–47.
29. Hammond CJ, Snieder H, Spector TD, Gilbert CE. Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. N Engl J Med 2000;342:1786–90.
30. Liao J, Su X, Chen P, et al. Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract. Hum Mol Genet 2014;23:6119–28.
31. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. Int J Epidemiol 2013;42:76–85.
32. Teucher B, Skinner J, Skidmore PM, et al. Dietary patterns and heritability of food choice in a UK female twin cohort. Twin Res Hum Genet 2007;10:734–48.
33. Bingham SA, Welch AA, McTaggart A, et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. Public Health Nutr 2001;4:847–58.
34. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:17–27.
35. McCance RA, Holland B, Widdowson EM, et al. McCance and Widdowson's The Composition of Foods. 5th revision and extended edition. Cambridge, UK: Royal Society of Chemistry; 1991:xiii, 462.
36. Christensen K, Gaist D, Vaupel JW, McGue M. Genetic contribution to rate of change in functional abilities among

Danish twins aged 75 years or more. Am J Epidemiol 2002;155:132–9.
37. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. Am J Respir Crit Care Med 2001;164:1655–9.
38. Steves CJ, Jackson SH, Spector TD. Cognitive change in older women using a computerised battery: a longitudinal quantitative genetic twin study. Behav Genet 2013;43:468–79.
39. Beebe DC, Holekamp NM, Shui YB. Oxidative damage and the prevention of age-related cataracts. Ophthalmic Res 2010;44:155–65.
40. Shui YB, Holekamp NM, Kramer BC, et al. The gel state of the vitreous and ascorbate-dependent oxygen consumption: relationship to the etiology of nuclear cataracts. Arch Ophthalmol 2009;127:475–82.
41. Reddy VN, Kasahara E, Hiraoka M, et al. Effects of variation in superoxide dismutases (SOD) on oxidative stress and apoptosis in lens epithelium. Exp Eye Res 2004;79:859–68.
42. Tarwadi K, Agte V. Linkages of antioxidant, micronutrient, and socioeconomic status with the degree of oxidative stress and lens opacity in Indian cataract patients. Nutrition 2004;20:261–7.
43. Cekic O, Bardak Y, Totan Y, et al. Nickel, chromium, manganese, iron and aluminum levels in human cataractous and normal lenses. Ophthalmic Res 1999;31:332–6.
44. Racz P, Ordogh M. Investigations on trace elements in normal and senile cataractous lenses. Activation analysis of copper, zinc, manganese, cobalt, rubidium, scandium, and nickel. Albrecht Von Graefes Arch Klin Exp Ophthalmol 1977;204:67–72.
45. Kuzniarz M, Mitchell P, Cumming RG, Flood VM. Use of vitamin supplements and cataract: the Blue Mountains Eye Study. Am J Ophthalmol 2001;132:19–26.
46. Chylack LT Jr, Wolfe JK, Singer DM, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. Arch Ophthalmol 1993;111:831–6.
47. Andrew T, Hart DJ, Snieder H, et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res 2001;4:464–77.

Footnotes and Financial Disclosures

Originally received: September 29, 2015.
Final revision: January 13, 2016.
Accepted: January 24, 2016.
Available online: March 23, 2016. Manuscript no. 2015-1709.

[1] Department of Twin Research and Genetic Epidemiology, Kings College London, London, United Kingdom.

[2] Department of Ophthalmology, Kings College London, London, United Kingdom.

[3] University of Warwick Medical School, Coventry, United Kingdom.

[4] London School of Hygiene and Tropical Medicine, London, United Kingdom.

Presented in part at the Association for Research in Vision and Ophthalmology Meeting, May 6–10, 2012, Fort Lauderdale, Florida.

Financial Disclosure(s):
The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Funded by the Wellcome Trust and the Guide Dogs for the Blind Association. The Twins UK study was also funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also received support from the National Institute for Health Research–funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London.

E.Y.-D.: Funding – Biotechnology and Biological Sciences Research Council and the London Interdisciplinary Doctoral Program.

P.G.H.: Recipient – Fight for Sight ECI award.

K.M.W.: Support – Medical Research Council Clinical Research Training Fellowship. The sponsors or funding organizations had no role in the design or conduct of this research.

Author Contributions:

Conception and design: Yonova-Doing, Forkin, Gilbert, Hammond

Data collection: Yonova-Doing, Forkin, Hysi, Williams, Spector

Analysis and interpretation: Yonova-Doing, Forkin, Hysi, Williams, Spector

Obtained funding: Not applicable

Overall responsibility: Yonova-Doing, Forkin, Gilbert, Hammond

Abbreviations and Acronyms:
CI = confidence interval; **DZ** = dizygotic; **FFQ** = food frequency questionnaire; **HATS** = Healthy Ageing in Twins; **LOCS** = Lens Opacity Classification System; **MZ** = monozygotic; **NDS** = nuclear dip score; **RRR** = relative risk ratio; **SD** = standard deviation.

Correspondence:
Christopher J. Hammond, MD, FRCOphth, Departments of Ophthalmology & Twin Research, King's College London, 3rd Floor, Block D, South Wing, St. Thomas' Hospital, Westminster Bridge Rd., London SE1 7EH, UK. E-mail: chris.hammond@kcl.ac.uk.